

DOCUMENT RESUME

ED 078 843

LI 004 378

AUTHOR Young, Carol Elizabeth
TITLE Development of Language Analysis Procedures With
Application to Automatic Indexing.
INSTITUTION Ohio State Univ., Columbus. Computer and Information
Science Research Center.
SPONS AGENCY National Science Foundation, Washington, D.C. Office
of Science Information Services.
REPORT NO OSU-CISRC-TR-73-2
PUB DATE Apr 73
NOTE 314p.; (25 References)
EDRS PRICE MF-\$0.65 HC-\$13.16
DESCRIPTORS *Automatic Indexing; *Computer Programs; *Electronic
Data Processing; Information Processing; *Information
Retrieval; *Language Patterns

ABSTRACT

Specifically, the problem which this research addresses is the development of procedures which define the relational attributes to words in English text. These relational attributes are considered an essential part of any good index and therefore an antecedent to the production of good indexes by automated means. The organization of this dissertation is as follows: Chapter I presents an overview of theories of language. Chapter II presents a theoretical framework for an empirical investigation of language. Chapter III presents procedures for the identification of relational attributes among words in English text. Chapter IV describes procedures for the identification and characterization, in relational terms, of clauses and phrases in text. Chapter V presents procedures for the assignment of case roles. These case roles amount to functional interpretation of text elements (e.g., phrases). Chapter VI relates the language analysis procedures presented in earlier chapters to the notion of indexing and proposes a graphical representation of English text as a general base from which a variety of indexes may be automatically derived. Chapter VII contains a brief statement of conclusions and of direction for future research. A KWIC index of the references cited is included in Appendix F.
(Author/NH)

ED 078843

LI 004 378

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

(OSU-CISRC-TR-73-2)

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE-
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

DEVELOPMENT OF LANGUAGE ANALYSIS PROCEDURES
WITH APPLICATION TO AUTOMATIC INDEXING

by

Carol Elizabeth Young

Work performed under
Grant No. 534.1, National Science Foundation

Computer and Information Science Research Center
The Ohio State University
Columbus, Ohio 43210
April 1973

FILMED FROM BEST AVAILABLE COPY

PREFACE

This work was done in partial fulfillment of the requirements for a doctor of philosophy degree in Computer and Information Science from The Ohio State University. It was supported in part by Grant No. GN 534.1 from the Office of Science Information Service, National Science Foundation, to the Computer and Information Science Research Center of The Ohio State University.

The Computer and Information Science Research Center of The Ohio State University is an interdisciplinary research organization which consists of the staff, graduate students, and faculty of many University departments and laboratories. This report is based on research accomplished in cooperation with the Department of Computer and Information Science.

The research was administered and monitored by The Ohio State University Research Foundation.

ACKNOWLEDGEMENTS

I wish to thank my adviser, Dr. James E. Rush, for his help with the preparation of this work and for the invaluable educational experience which he has provided during my graduate career. I am grateful to the other members of my reading committee, Dr. H. William Buttelmann, Dr. Ronald L. Ernst and Dr. Anthony E. Petrarca for their contributions to this work. I am grateful to Dr. Bertrand Clovis Landry and Dr. Harold B. Pepinsky for reading the drafts of this dissertation and for providing valuable suggestions. Many of the ideas contained in this dissertation were a result of the interaction of a group of computer scientists, linguists and psychologists. This group consists of Dr. Walter A. Cook, S.J., Dr. Bertrand C. Landry, Dr. Naomi M. Meara, Dr. Harold B. Pepinsky, Mrs. Suzanne M. Strong and Father John A. Valley. I would especially like to thank Dr. Landry and Dr. Meara for their contribution in developing the computer programs discussed in Chapters IV and V. I am indebted to Mrs. Mary Kimball for her help in transforming the manuscript of this dissertation to typed copy.

I wish to thank my parents and Mr. Leo Scheer for their moral support during my graduate career. I extend my gratitude for the encouragement and the financial assistance of Mr. D. Balleé, who made the completion of this work possible.

During my graduate studies, I received financial aid from a University Fellowship and research support from a grant (GN-534.1) from the National Science Foundation to the Computer and Information Science Research Center. The Mershon Center Programs of Research in Education Leadership and Public Policy provided partial support for the preparation of this dissertation. Computer time for this research was provided by the grant and The Ohio State University Instruction and Research Computer Center, Roy Reeves, Director.

TABLE OF CONTENTS

	<u>Page</u>
PREFACE	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	x
LIST OF FIGURES	xii
QUOTATION REFERENCES	xiv
STATEMENT OF THE PROBLEM	xv
 CHAPTER I. OVERVIEW OF THEORIES OF LANGUAGE	 1
1. Introduction	1
2. Traditional (Pedagogical) Grammar	3
3. Contemporary Approaches to a Grammar of English	5
References	8
 CHAPTER II. A THEORETICAL FRAMEWORK FOR AN EMPIRICAL INVESTIGATION OF LANGUAGE	 9
1. Introduction	9
2. Function and Purpose of Theory	10
3. A Theoretical Framework for the Description and Analysis of Written English	12
3.1 Introduction	12
3.2 The Basic Elements of Language	12
3.3 The Naming Process in Language	15
3.3.1. Classification of NAMES	16
3.3.2. Classification of RELATIONS	18
3.3.3. Definition of "Sentence"	20
3.3.4. Interregnum	20
3.4. Formal Statement of a Theory of Language	21
3.4.1. PRIMARY RELATIONS	21
3.4.2. SECONDARY RELATIONS	29
3.4.2.1. The CONJUNCTIVE SECONDARY RELATIONS	29
3.4.2.2. The ATTRIBUTIVE SECONDARY RELATIONS	31
3.4.3. NAMES	32
3.4.3.1. The Notion of PHRASE	32
3.4.3.2. The Notion of CLAUSE	32
3.4.4. SENTENCE	36
References	47

	<u>Page</u>
CHAPTER III. IDENTIFICATION OF SIMPLE NAMES AND RELATIONS . . .	49
1. Introduction	49
2. Overview of Related Work in Syntactic Analysis	49
2.1. Review of Syntactic Analysis Procedures	
--Part I	50
2.1.1. The Multiple-Path Syntactic Analyzer	51
2.1.2. Syntactic Analysis for Transformational Grammar	52
2.1.3. Syntactic Analysis Based on String Analysis	53
2.1.4. Syntactic Analysis for the Simulation of Natural Language Processes	54
2.2. Review of Syntactic Analysis Procedures	
--Part II	55
2.2.1. The Economical Parser	55
2.2.2. The Computational Grammar Coder	56
2.2.3. The WISSYN-System	59
2.2.4. Syntactic Analysis Based on a Regular Grammar	61
2.2.5. A Limited Program for Syntactic Analysis.	62
2.2.6. Syntactic Analysis Based on Transition Network Grammars	63
2.3. Summary	64
3. A Basis for Automated Syntactic Analysis	64
3.1. Defining Grammatical Classes	65
3.2. Formal Definition of Function Words	70
3.2.1. Definition of Function Words	70
3.2.2. Rules for Determination of Grammatical Classes	78
3.2.2.1. Rules for the Class DTR	78
3.2.2.2. Rules for the Class AMB	81
3.2.2.3. Rules for the Class PRN	81
3.2.2.4. Rules for the Class INT	82
3.2.2.5. Rules for the Class REL	82
3.2.2.6. The Class ADV	84
3.2.2.7. Rules for the Classes AJN, AUX and MOD	84
3.2.2.8. The Class VRB	90
3.2.2.9. Rules for the Classes CCN, CCP and SCN	90
3.2.2.10. Rules for the Class THR	91
3.2.2.11. Rules for the Class THT	92
3.2.2.12. Rules for the Class PRP	92

	<u>Page</u>
3.2.2.13. Rules for the Class NEV	94
3.2.2.14. Rules for the Class NEG	94
3.2.3. Sufficiency of the Rules	95
4. MYRA - A Program for Grammatical Class Assignment	98
4.1. General Description of MYRA	98
4.1.1. The Dictionary	98
4.1.2. The Rules	99
4.2. Operation of MYRA	99
4.2.1. Dictionary Look-Up	101
4.2.2. Application of Rules 1-101	102
4.2.3. Application of Rules 102-101	103
5. Experimental Testing of MYRA	104
6. Summary	107
References	108
 CHAPTER IV. IDENTIFICATION OF COMPOSITE AND COMPLEX NAMES	 111
1. Introduction	111
2. Procedures for the Identification of CLAUSES	113
2.1. Phase I of CAP/I	113
2.1.1. Rules for the Class CCN	114
2.1.2. Rules for the Class REL	114
2.1.3. Rules for the Class THT	114
2.1.4. Rules for the Class PNT	115
2.2. Phase Two of CAP/I	115
2.2.1. Rule for the Class SCN	115
2.2.2. Rule for the Class INF	115
2.2.3. Rule for the Class PTC	115
2.3. Experimental Results from CAP/I	116
2.3.1. Identification and Analysis of CAP/I Errors	 116
2.3.2. Conclusions From Analysis of CAP/I Results	 118
3. Procedures for the Identification of PHRASES (PAP)	118
3.1. Procedures for the Reduction of Certain COMPLEX NAMES to SIMPLE NAMES	 120
3.2. Procedures for the Identification of PHRASES	121
3.3. Procedures for the Reduction of Certain COMPLEX NAMES to PHRASES	 124
3.3.1. Reductions Not Involving CCP	125
3.4. Results of Tests	126
4. Improved Procedures for the Identification of CLAUSES	131
4.1. Rules for Identifying CLAUSES which Involve the Class INF	 131
4.2. Rules for Identifying CLAUSES which Involve the Class PTC	 132

	<u>Page</u>
4.3. Other CLAUSE Identification Rules	133
References	135
 CHAPTER V. THE FUNCTION OF NAMES AND RELATIONS	 136
1. Introduction	136
2. Case Grammar	136
2.1. The Basic Components of Case Grammar	137
2.2. Categories of PRIMARY RELATION	138
2.3. Case Roles	139
3. Identification of Case Roles	142
3.1. Essential Case Roles	142
3.2. Peripheral Case Roles	145
3.3. Assignment of Case Roles to CLAUSES	147
4. The Case Grammar Program, CGP	147
5. Experimental Testing of CGP	148
6. Summary	151
References	153
 CHAPTER VI. LANGUAGE ANALYSIS AND INDEXING	 155
1. Introduction	155
2. Indexes and Indexing	155
3. Indexing Theory	158
4. Procedures for the Graphic Representation of Sentences	 160
4.1. Why a Graphic Representation?	161
4.2. Approaches to the Graphic Representation of Language	 162
4.2.1. Graphic Structure Proposed by Salton	162
4.2.2. TOSAR Graphs	162
4.2.3. Graphic Representation of Deep Structure	164
4.2.4. Graphic Structures Proposed by Plath	166
4.2.5. Graphic Structures Proposed by Tesniere	169
4.3. A Proposal for the Production of Graphic Representations of English Sentences	 169
4.3.1. Definition of Components	169
4.3.2. Representation of PRIMARY-RELATION Nodes	 171
4.3.3. Representation of Case Types and COMPOSITE NAMES	 171
4.3.4. Representation of Consecutive SECONDARY PHRASES	 174

	<u>Page</u>
4.3.5. Representation of PHRASEs Related by COORDINATE CONJUNCTIVE SECONDARY RELATIONS	176
4.3.6. Representation of CLAUSES Joined by CCN	179
4.3.7. Representation of CLAUSES Joined by SCN	181
4.3.8. Representation of NOMINAL CLAUSES . . .	182
4.3.9. Representation of ADJECTIVAL CLAUSES .	186
4.3.10. Representation of ADVERBIAL CLAUSES . .	188
5. Summary	189
References	191
 CHAPTER VII. CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH . .	 193
1. Conclusions	193
 APPENDIX A	 194
References	204
 APPENDIX B	 205
 APPENDIX C	 216
 APPENDIX D	 234
 APPENDIX E	 258
 APPENDIX F	 271

LIST OF TABLES

		<u>Page</u>
2.1	Examples of SIMPLE, COMPOSITE and COMPLEX NAMES from English.	17
2.2	Summary of Defined Terms	22
2.3	Classes of ALLIED RELATION and their Elements.	27
2.4	RELATIONS ending in "ing" which are not PARTICIPIAL RM-RELATIONS.	28
2.5	Examples of NAMES, PHRASES and CLAUSES	35
2.6	Definition of Prefix-Notation Symbols as Used in the Specification of SENTENCES.	38
2.7	Specification of SENTENCES.	39
3.1	Specification of the Elements of Each Terminal Class in SECONDARY RELATION.	73
3.2	Specification of the Elements of each Class of PRIMARY RELATION which are Elements of FUNCTION WORD.	74
3.3	Specification of the Elements of NAME which are Elements of FUNCTION WORD.	76
3.4	Composition of the Class of FUNCTION WORDS.	77
3.5	Definition of Notation Symbols Used in the Rules for Grammatical Class Assignment.	79
3.6	Error Analysis of the Results Produced by MYRA Using the Limited Dictionary	105
3.7	Error Analysis of the Results Produced by MYRA Using the Extended Dictionary	105
4.1	Accuracy of the Output of CAP/I	119
4.2	Rules for the Identification and Characterization of PHRASES.	122
4.3	Accuracy of the Output of PAP	127

	<u>Page</u>
5.1 Case Roles and their Definitions According to Fillmore	140
5.2 Definition of Essential and Peripheral-Case Roles as Used in the Present Research	141
5.3 Case Roles as after Fillmore and as Used in this Research	143
5.4 Case-Role Assignments for Sample CLAUSES each Containing One of the Categories of PRIMARY PHRASE	144
5.5 Case Frames for Essential Cases as Prescribed by the PRIMARY PHRASE Category	146
5.6 Accuracy Attained in Case Grammar Assignment Made by CGP	152
6.1 Symbols Used to Represent Case Grammar Classes of PRIMARY RELATIONS	172
6.2 The Case Roles Along with the Corresponding Numerical Codes Used in the Graphic Representation	173
6.3 COORDINATE CONJUNCTIVE SECONDARY RELATIONS and their Corresponding Numerical Codes.	177

LIST OF FIGURES

	<u>Page</u>
1.1 Illustrations of various concepts which comprise the notion of "language"	2
2.1 A hierarchical division of the types of RELATIONS	19
2.2 An example of a sentence analyzed in terms of NAMES and RELATIONS	24
2.3 Relationships between categories of RELATION	25
2.4 The complete hierarchy of PRIMARY RELATIONS	30
2.5 Hierarchical partitioning of SECONDARY RELATION	33
2.6 Partitioning of the class of PHRASEs	34
2.7 Partitioning of the class of CLAUSES	34
3.1 Complete hierarchical division of SECONDARY RELATIONS	72
3.2 Hierarchy of NAMES showing particularly classes which are numbers of FUNCTION WORD.	75
3.3 A distribution of dictionary elements by length	100
3.4 Sample output produced by MYRA from the text <u>The Old Man and the Sea</u>	106
4.1 Sample output produced by CAP/I from <u>The Old Man and the Sea</u>	117
4.2 Sample output produced by PAP from <u>The Old Man and the Sea</u>	128
4.3 The operation of the language analysis procedures	129
4.4 The proposed operation of the language analysis procedures	130
5.1 A flow diagram of CGP	149
5.2 Sample output produced by CGP from <u>The Old Man and the Sea</u>	150

	<u>Page</u>
6.1 Landry's conceptualization of an index space	159
6.2 Graphic representation proposed by Salton	163
6.3 A graph produced via the TOSAR model of the sentence: Oligomerization of propylene with the aid of Al(Alkyl)_3 to obtain hexene, and separation and purification of the excess propylene by fractional distillation and recycling of the propylene	165
6.4 The "underlying meaning" of the sentence "John's love is good" as represented by the Conceptual Dependency Parser	167
6.5 Plath's automatically produced sentence diagrams	168
6.6 Tesnierre analysis of the sentence "Le feu vert indique la voie libre" (The green light indicates clear passage)	170

QUOTATION REFERENCES

Page

- 1 R. W. Langacker, Fundamentals of Linguistic Analysis,
Harcourt Brace Jovanovich, Inc., New York, 1972,
p. 1.
- 1 J. E. Hopcroft and J. D. Ullman, Formal Languages and
their Relation to Automata, Addison-Wesley Publishing
Co., Massachusetts, 1969, p. 1.
- 1 A. Bierce, The Enlarged Devils' Dictionary, Doubleday
and Company, Inc., New York, 1967, p. 80.
- 3 J. Moffat, The Structure of English, Pegasus, Christchurch,
England, 1968, p. 28.
- 5 N. Chomsky, Aspects of the Theory of Syntax, Cambridge:
M.I.T. Press, 1965, p. 15.
- 5 Hopcroft and Ullman, op. cit., p. 16.
- 9 T. H. Huxley, Darwiniana, D. Appleton-Century Co., Inc.,
New York, New York, 1897, pp. 368-369.
- 15 L. Carroll, Through the Looking Glass, Random House,
New York, New York, 1946, p. 148.
- 49 A. E. Nichols, English Syntax, Holt, Rinehart and Winston,
New York, New York, 1965, p. 3.
- 111 Sir James Jeans, The Universe Around Us, 4th edition,
The University Press, Cambridge, England, 1945.
- 111 Nichols, op. cit., p. 54.
- 155 J. Gardin, "Semantic Analysis Procedures in the Sciences
of Man," Social Science Information 8(1), 34, 1969.

STATEMENT OF THE PROBLEM

The problem addressed in this research is one that has received the attention of many people. It is this: "Is it possible to develop computer programs which operate upon written English so as to produce a form of representation that makes possible the generation of high quality indexes by automated means?" Automated indexing procedures have been largely relegated to the tasks of sorting, formatting and printing, while the substance of the index is manually derived. Thus the question is one of how to derive good, meaningful indexes from documents automatically rather than manually.

Specifically, the problem which this research addresses is the development of procedures which define the relational attributes to words in English text. These relational attributes are considered an essential part of any good index and therefore an antecedent to the production of good indexes by automated means.

The organization of this dissertation is as follows:

Chapter I presents an overview of theories of language.

Chapter II presents a theoretical framework for an empirical investigation of language.

Chapter III presents procedures for the identification of relational attributes among words in English text.

Chapter IV describes procedures for the identification and characterization, in relational terms, of clauses and phrases in text.

Chapter V presents procedures for the assignment of case roles. These case roles amount to functional interpretation of text elements (e.g., phrases).

Chapter VI relates the language analysis procedures presented in earlier chapters to the notion of indexing and proposes a graphical representation of English text as a general base from which a variety of indexes may be automatically derived.

Chapter VII contains a brief statement of conclusions and of direction for future research.

Finally, a KWIC index of all the references cited in this dissertation is included as a way of making parts of the dissertation accessible by other than sequential means. This index is given in Appendix F.

CHAPTER I. OVERVIEW OF THEORIES OF LANGUAGE

A language is a set of principles relating meanings and phonetic sequences.

R. W. Langacker, Fundamentals of Linguistic Analysis

A language is any set of sentences over an alphabet. A sentence over an alphabet is any string of finite length composed of symbols from the alphabet. An alphabet or vocabulary is any finite set of symbols.

J. E. Hopcroft and J. D. Ullman, Formal Languages and their Relation to Automata

English, n., A language so haughty and reserved that few writers succeed in getting on terms of familiarity with it.

A. Bierce, The Enlarged Devil's Dictionary

1. Introduction

Printed or written language consists of a set of elements called words.

This set of words is the vocabulary of the language. A dictionary is a collection of vocabulary elements along with a description of each element. The permissible ways in which vocabulary elements may be strung together (i.e., arranged in a linear sequence) is governed by a set of rules called a grammar. The resulting sequence may be called a sentence. The set of sentences permitted by the grammar constitute a language (1), (see Figure 1.1).

While the derivation and definition of words is of interest, especially to the lexicographer, etymologist, historical linguist and sociolinguist, only a small part of modern linguistic effort is directed toward the study of words. Most attention has been paid to the grammar,

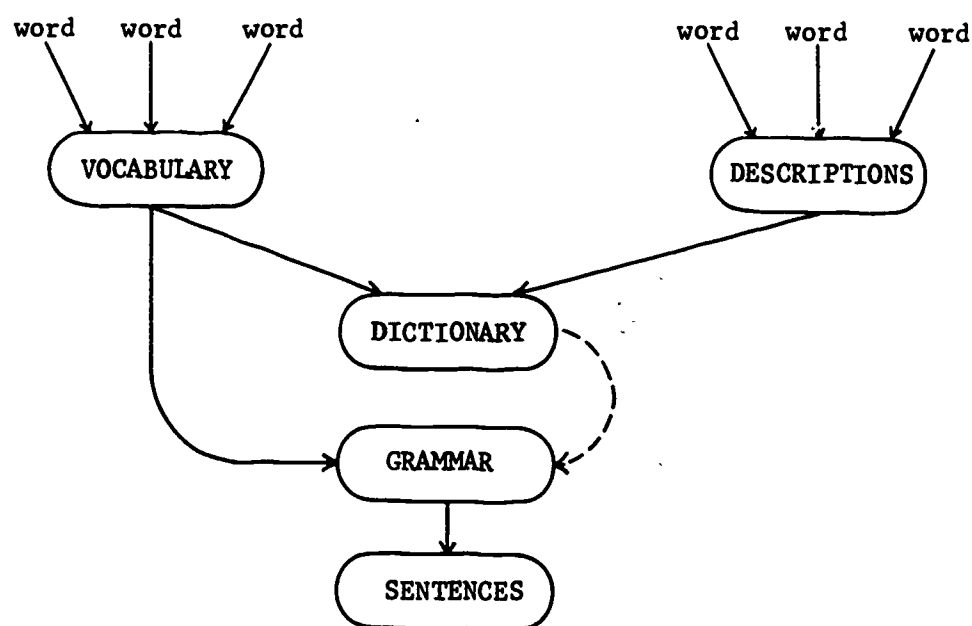


Figure 1.1 Illustration of various concepts which comprise the notion of "language."

or grammars, of a language, for the most part because the grammars of natural languages are still largely intuitive and difficult to describe precisely.

2. Traditional (Pedagogical) Grammar

Grammar describes the relationships which words bear to one another in a statement.

J. Moffat, The Structure of English

Traditional grammar has been viewed as a model which dictates the form of language. A sentence construction was, within the context of this model either right or wrong, and grammar books contain the last word on the validity of a string of words as a sentence. The traditional grammar model has been handed down for centuries as something not to be questioned. Let us consider how this model came into being.

The study of language can be traced back to the Greeks of the 5th century B.C., when language was studied from philosophical perspectives. Language was believed to contain the universal forms of thought. Syntactic forms of the language were derived and were defined in accordance with the philosophical significance of each form. The Greek theories were developed through an investigation of the processes governing thought and action (2).

The parallel between this model of grammar which developed at such an early period and our current traditional grammar is great. Here are some of the steps in the development of Greek grammar. Plato, in one of his dialogues, classified nouns and verbs on logical grounds. Aristotle identified nouns, verbs and conjunctions (although the use of

the latter term differs from its current use). The Stoics added a class of articles and examined number, voice, mood and case (nominative, vocative, accusative, genitive and dative). Dionysios Thrax, an Alexandrian scholar of the 1st century B.C., recognized eight parts of speech: noun, verb, participle, article, pronoun, preposition, adverb and conjunction. Remmius Palaemon added the interjection as a part of speech in the first century A.D. This classical grammar survived in over a thousand manuscripts and it formed the basis of the Latin grammar recorded in the eighteenth century (3).

In the process of writing down an English grammar, early grammarians were influenced by logic and by Latin grammar. To scholars of the 18th century a correct sentence had to be logically correct. Thus, a rule was put forth that words like "perfect," "round" and "square" could not have a comparative degree. If a thing is perfect, it can not be more perfect since one cannot reach beyond perfection. If a thing is round, it is already round and cannot become rounder. A modern linguist would agree neither with the initial hypothesis nor with the outcome.

Latin grammar was a major factor in the formation of an English grammar. The transference of Latin rules to English rested on the theory that all languages had a common structure. As a result, traditional grammar books refer to accusative and dative cases when there is no need for them in English; to gender in nouns when gender in English nouns is a part of their meaning; and to the restriction from placing a preposition at the end of a sentence when this is not always

possible (4)¹. While such attributes are important (e.g., in the investigation of the derivation of English from Latin), these attributes are not relevant to the definition of a grammar of English.

3. Contemporary Approaches to a Grammar of English

The grammar of a language is a system of rules that determine a certain pairing of sound and meaning. It consists of a syntactic component, a semantic component and a phonological component.

N. Chomsky, Aspects of the Theory of Syntax

Formally, we denote a grammar G by (V_N, V_T, P, S) . The symbols V_N , V_T , P and S are, respectively, the variables, terminals, productions and start symbol.

J. E. Hopcroft and J. D. Ullman, Formal Languages and their Relation to Automata

While traditional grammar has apparently been adequate for pedagogical purposes, its inadequacies for other purposes became acutely apparent in the 1950's as a consequence of the futile attempts to automatically translate from one natural language to another. The main result of the vast amounts of money, time and effort that these attempts consumed was the conclusion that much more research was needed to examine basic structures and operators of language (5).

Since that time, several theories of language have been developed, including theories of formal language, using the principles of propositional calculus, group theory and automata theory. These theories,

1. The frequent inappropriateness of this rule is illustrated by a statement attributed to Winston Churchill: "That is an imposition up with which I will not put!"

although often readily susceptible to implementation,² deal primarily with restricted subsets of English or with "artificial" languages and are therefore not applicable to English on a broad scale. Other theories, which are descriptive in nature and are broadly applicable to natural language have not been easy to implement because of the complexity of the procedures or because of the large size of the vocabulary of a language. Grammars based on theories in this category include the transformational-generative (6), dependency (7), immediate constituent (8), phrase structure (9), predictive (10), systemic (11) and stratified (12) grammars of language. Versions of some of these grammars have been implemented, and have provided valuable empirical data. However, their implementation on a practical scale, (e.g., for use with a large data base of technical documents where the time to handle a single sentence is important) seems impossible. The severity of the time limitation suggested here can be seen in an attempt by Winograd to implement a theory of language. While his work in this area is extensive, and his results important, the system he has developed deals with an extremely limited universe of discourse (13), and its extension to some more extensive universe of discourse seems impractical.

All of the theories mentioned here have unquestionably contributed to our understanding of language, but I believe they fail to account for the basic purpose of language and that they therefore fail to deal

2. Everywhere in this dissertation the word implement (and its derivatives) means "to render in a form suitable for processing by a digital computer."

with the fundamental properties of language. In the next chapter, I propose a theoretical framework for studying the English language. I believe this theory makes it possible to operate on English utterances in algorithmic terms and that the results of these operations will be useful in developing automated indexing procedures.

References

1. S. Ginsberg, "Lectures on Context-Free Languages," in M. A. Arbib (ed.), Algebraic Theory of Machines, Languages, and Semigroups, Academic Press, New York, 1968, pp. 313-327.
2. L. Bloomfield, Language, Holt, Rinehart and Winston, New York, 1933, p. 564.
3. J. Moffat, The Structure of English, Pegasus Press, Christchurch, New Zealand, 1968, p. 135.
4. N. R. Cattell, The New English Grammar: A Descriptive Introduction, The MIT Press, Cambridge, Massachusetts, 1966, p. 162.
5. Y. Bar-Hillel, "Four Lectures on Algebraic Linguistics and Machine Translation," in Automatic Translation of Languages, Pergamon Press, New York, 1966, pp. 1-27.
6. N. Chomsky, Aspects of the Theory of Syntax, MIT Press, Cambridge, Massachusetts, 1965, p. 251.
7. D. G. Hayes, "Dependency Theory: A Formalism and Some Observations," Language 40 (04), 511-524 (1964).
8. J. Lyons, Introduction to Theoretical Linguistics, University Press, Cambridge, 1968, p. 209.
9. ibid., p. 215.
10. A. Oettinger and S. Kuno, "Multiple-Path Syntactic Analyzer," Proceedings of the International Federation of Information Processing Congress, Munich, 1962.
11. M. A. K. Halliday, "Notes on Transitivity and Theme in English, Part 1", Journal of Linguistics 3(1), 37-81 (1967)
12. S. M. Lamb, Outline of Stratification Grammar, University of California, Berkeley, 1962, pp. 64.
13. T. Winograd, "Understanding Natural Language," Cognitive Psychology 3, 1-191 (1970).

CHAPTER II. A THEORETICAL FRAMEWORK FOR AN EMPIRICAL INVESTIGATION OF LANGUAGE

I want to put the case clearly before you, and I will therefore show you what I mean by another familiar example. I will suppose that one of you, on coming down in the morning to the parlour of your house, finds that a tea-pot and some spoons which had been left in the room on the previous evening are gone,--the window is open, and you observe the mark of a dirty hand on the window-frame, and perhaps, in addition to that, you notice the impress of a hob-nailed shoe on the gravel outside. All these phenomena have struck your attention instantly, and before two seconds have passed you say, "Oh, somebody has broken open the window, entered the room, and run off with the spoons and the tea-pot!" That speech is out of your mouth in a moment. And you will probably add, "I know there has; I am quite sure of it!" You mean to say exactly what you know; but in reality you are giving expression to what is, in all essential particulars, an hypothesis. You do not know it at all; it is nothing but an hypothesis rapidly framed in your own mind. And it is an hypothesis founded on a long train of inductions and deductions.

T. H. Huxley, Darwiniana

1. Introduction

A theoretical framework is essential to begin an experimental work. An hypothesis may be changed or discarded or new hypotheses may emerge in the course of an investigation, but only within a theoretical framework can experimental results assume general implications.

A theory rests upon observation and is verified through observation. Thus in any study of language one must gather evidence which may enable him to hypothesize about causes and effects. When we ask questions such as "What is the function of the preposition 'of' in English?" it is clear that we are already pretty far down a particular line of investigation having, apparently, observed such things as words, classified them (among other things) according to some scheme, and determined that at

least some words have particular functions which it is our desire to determine.

While it is desirable that theories be as general as possible, it is rare that general theories emerge at early stages of an investigation. Rather theories are often devised for particular purposes and without any particular concern for generality. That is the case with the theoretical framework of English described in this Chapter. Prior to describing this theory, however, I feel it is necessary to make clear to the reader my views on the function and purpose of a theory. I attempt to do this in the following section. I hasten to point out that the remarks presented below are not to be construed as expressing any original thought on my part. On the contrary, these views are acknowledged to be the basis of scientific investigation (1). But scientists and laymen alike sometimes become so convinced of a theory, that the theory is accepted as a law. This acceptance has been the case in the treatment of the traditional model of grammar as an absolute. I therefore feel it is necessary to emphasize what Huxley has so admirably expressed in the quotation at the beginning of this Chapter concerning the nature of theory.

2. Function and Purpose of Theory

A theory is a way of representing, organizing and observing phenomena. It is a statement about the way in which one views and perceives phenomena (2). A theory is not purported to be an absolute. At one time, people held to the theory that the universe revolved around the earth. Ptolemy, based on this view of the universe, developed a system

of epicycles to describe the movement of the universe. The mathematical formulations which he developed precisely described and predicted the movement of the planets. The theory is internally consistent and, even more impressive, experimentally verifiable. Now that man can look out from the stars, his view of the universe has changed and has necessitated different theories. But this does not destroy the consistency or precision of the earlier theory. And there is no guarantee, that at some future date, a new dimension will be added to the observables we now comprehend, and new theories of the behavior of the universe will emerge. This example illustrates that the structure set down by a theory and the relationships postulated among the objects with which it deals is not an attribute of the objects themselves. On the contrary, a theory imposes some form of logic or structure upon observables and provides for the interpretation of the observables within this framework (3).

For my purposes, a theory must not only be adequate in some abstract sense, it must also be useful. But how is utility to be measured? I can only say that for me a theory is useful so long as it helps toward the achievement of certain goals I have established for myself. Of course goals have a way of changing with time and thus so must theories. In effect, then, I judge a theory according to whether it serves me well in some investigation.

The main purpose of this chapter is to describe a theoretical framework for language which will serve for the analysis and description of language, at least of English, in terms that are considered beneficial in indexing, especially automatic indexing. This theoretical framework

has permitted the expression of rules upon which algorithms³ have been built to effect various analyses of English. These algorithms are described in Chapters III-V.

3. A Theoretical Framework for the Description and Analysis of Written English

3.1. Introduction

Language is a means of communication. But this sweeping statement is not particularly helpful. However, if one accepts English as a subset of language the subject matter to be dealt with is reduced to somewhat more manageable proportions, thereby excluding body language, chemistry, mathematics, and other languages from consideration. But English may be written or spoken and it seems that the properties of the two are sufficiently different that one may further reduce the subject matter by limiting one's attention to written English. In particular, the language dealt with in this research is that of technical and literary works.

3.2. The Basic Elements of Language

Whatever language is being considered, it seems to me that it may be described as a system of things and relations between them. In general, what are observed day by day are things⁴. Furthermore, these

3. An algorithm is a set of precise, unambiguous rules the application of which must produce results that are independent of the machine (or person) applying them (4).

4. Simmons, et al. (5), use the term "concept" instead of thing, but I prefer "thing" as being potentially more general.

things exhibit changes with time. It is frequently found by observers to be useful or necessary to name the things and the changes they exhibit. For instance, when a rock is observed to fall from a cliff and when the observer wishes to communicate with someone about the changes the thing has undergone, the object may be called "rock" and the change in its location with respect to time "fall."⁵ If one is near the terminus of the rock's path, one might say the rock "dropped," whereas if one is near the point from which it fell, one might say the rock "fell." The place from which the rock came (or departed) may also be named, for example, "cliff." The important thing to note here is that names are given to things and to behaviors exhibited by (or, if you will, attributed to) the things. (Any observable change of state, no matter how slight, I shall call a behavior). It is also important to note that a great effort is made to differentiate between the two types of names. Thus, I argue that language has as its basic elements names of things and names of behaviors. But there is a fundamental difference between the role and nature of names of things and of names of behaviors. Since things are often directly sensible and behaviors are never sensible, the former are treated as though they were related by means of the latter. In the example above, of the falling rock, the rock is directly sensible (by touch, smell, etc.) whereas "falling" is a name given to the change in

5. I may surely be forgiven for using the language I am attempting to describe. Subsequently, a word or phrase given special meaning within the context of this dissertation will be printed in upper case letters.

location of the rock with respect to time. That is, "falling" is the name given to the relation between the rock and its successive locations in space at successive points in time. I hope this illustration serves to clarify the notion of name-of-things, which I shall call, simply, NAME and of name-of-behavior, which I shall call RELATION.

Many researchers have in one way or another treated languages as relational systems. Rothstein has proposed the use of binary relations in representing strings of a language (6, 7): In a model of verbal understanding, Simmons (8) has defined primitive elements of his model to be concepts and relations; these primitives are essentially equivalent to my thing and relation. And as will be seen subsequently, the views of Montgomery (9), Fillmore (10), Chafe (11), and others are compatible with the relational nature of English I put forward here. I must note that the relations of which I speak are of two types: the one having some referent in actual experience, the other being exclusively a linguistic device. This matter is discussed below in Section 3.3.2.

To summarize to this point, I have argued that a language consists of NAMES and RELATIONS. These names and relations are, in turn, strings of symbols formed from a basic set of symbols called an ALPHABET. Elementary alphabetical strings are called WORDS. A NAME is a word or string of words assigned to a thing, and a RELATION is a word or string of words assigned to a behavior. I have also said that a language provides for the production of more complex names or relations by combination of simpler names or relations. Let us now consider how this is done in English.

3.3. The Naming Process in Language

"The name of the song is called 'Haddocks' Eyes'." "Oh, that's the name of the song, is it?" Alice said, trying to feel interested. "No, you don't understand," the Knight said, looking a little vexed. "That's what the name is called. The name really is 'The Aged Aged Man.'" "Then I ought to have said 'That's what the song is called'?" Alice corrected herself. "No, you oughtn't: that's quite another thing! The song is called 'Ways and Means': but that's only what it's called, you know!" "Well, what is the song, then?" said Alice, who was by this time completely bewildered. "I was coming to that," the Knight said. "The song really is 'A-sitting On A Gate': and the tune's my own invention."

Lewis Carroll, Through the Looking Glass

While NAMES and RELATIONS denote basic elements of a language, it will usually be found that simple NAMES (single words) may be combined for naming many things or behaviors so that it is unnecessary to assign a unique name to every thing or behavior. In other words, the basic vocabulary⁶ of a language will be found usually to be rather quickly extended to practical limits (of memory, essentially) and that to continue to name things and behaviors in a practical way, additional linguistic devices are required. For instance, there are many horses in the world and it would be cumbersome at best to have to assign a unique word, as a NAME, to each and every one, even though the language might provide the capability of doing so. Instead languages provide for the modification of basic NAMES by permitting several NAMES to be related to one another in special ways. A horse may be brown or black or large or fast or wild or docile. It may run or walk or trot or gallop and these terms may be specified as fast or slow or fluid or jerky or stylish. My point is that by providing appropriately for combining simple NAMES to

6. Recall the definition of vocabulary given in Chapter I.

form more complex ones, a language becomes more powerful without the burden of a huge vocabulary. The relations between the NAMES may be explicitly represented or not. For example, the relation between "brown" and "horse" in the NAME "brown horse" is established, in English, by the positioning of the words, whereas explicit relational words are often used as in "house of wax," where "of" serves this purpose.

3.3.1. Classification of NAMES

A NAME may be simple, composite or complex. A SIMPLE NAME is a single word (string of alphabetical symbols exclusive of the blank) which is assigned to a thing. A COMPOSITE NAME is a consecutive sequence of SIMPLE NAMES separated by one or more blanks. A COMPLEX NAME is an ordered triple, $N_1 R N_2$, where N_1, N_2 are SIMPLE, COMPOSITE OR COMPLEX NAMES, or are vacuous, and R is a RELATION. Examples of these three classes of NAME are given in Table 2.1.

In order to establish a relation between the NAME types defined here and traditional linguistic terminology, the reader may note that, from the example of Table 2.1, SIMPLE NAMES seem to correspond with nouns, COMPOSITE NAMES with a noun preceded by a series of adjectives, and a COMPLEX NAME with no traditionally defined entity. It is emphasized that no effort has so far been made to investigate the function of NAMES in English, save that they represent things. Nor has any mention been made of possible methods for identifying them in English utterances. These matters are dealt with later.

Table 2.1 Examples of SIMPLE, COMPOSITE and COMPLEX NAMES from English.

SIMPLE	COMPOSITE	COMPLEX
horse	horse blanket	he is a man
trees	pine cone	a tree by the stream
rock	heavy black stone	rocks in his head
wind	cold front	the edge of a hurricane
water	running water	mountains shrouded in mist
ocean	white caps	
man	computer programs	Of Mice and Men
history	man's birth	
time	space/time continuum	
Mississippi	Missouri River	

3.3.2. Classification of Relations

RELATIONS are of two kinds: primary and secondary. A PRIMARY RELATION is defined as a single word which is assigned to a behavior. PRIMARY RELATIONS are, in turn, of two types: dominant and recessive. A DOMINANT PRIMARY RELATION (DP-RELATION) is a PRIMARY RELATION which can serve in isolation to relate one or more names. A RECESSIVE PRIMARY RELATION (RP-RELATION) is a PRIMARY RELATION which serves as an argument of another PRIMARY RELATION. SECONDARY RELATIONS are defined as special linguistic elements which have no referents in experimental behaviors but serve only to relate NAMES to NAMES or to relate RELATIONS to NAMES or to RELATIONS.

RELATIONS are also designated as simple or composite. A SIMPLE RELATION is a single PRIMARY or SECONDARY RELATION. A series of SECONDARY RELATIONS is a COMPOSITE SECONDARY RELATION (CS-RELATION). A series of DP-RELATIONS or a series of DP-RELATIONS followed by a SECONDARY RELATION is a COMPOSITE DOMINANT PRIMARY RELATION (CDP-RELATION). A series of RP-RELATIONS or a series of RP-RELATIONS followed by a SECONDARY RELATION is a COMPOSITE RECESSIVE PRIMARY RELATION (CRP-RELATION). An RP-RELATION in juxtaposition with a DP-RELATION cannot be subsumed within the DP-RELATION. Rather the RP-RELATION serves as a COMPLEX NAME or, as stated before, as an argument of the DP-RELATION.

This hierarchy of RELATIONS is shown in Figure 2.1. Relations have some connection with traditional linguistic terminology. PRIMARY RELATIONS correspond roughly with verbs, and SECONDARY RELATIONS with prepositions and conjunctions. RELATIONS in English are considered to

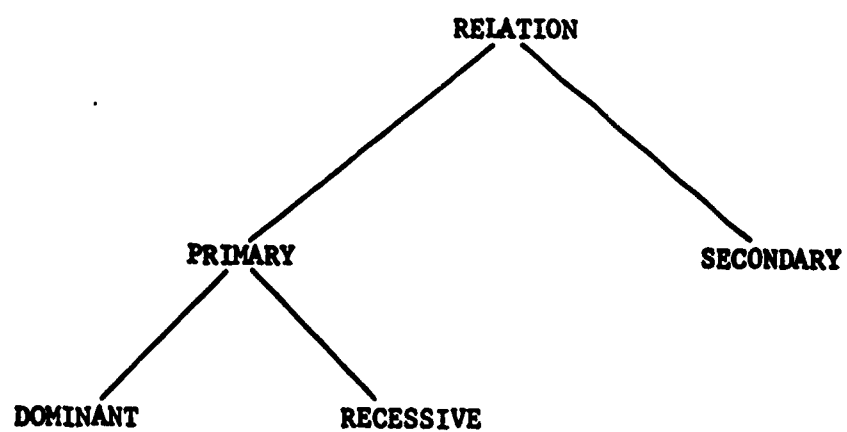


Figure 2.1 A hierarchical division of the types of RELATIONS.

function in the mathematical sense so that they take arguments just as do mathematical relations. It is important to note, too, that RELATIONS may connote various kinds of relationship between their arguments (such as equivalence, space/time, etc.). These connotations, though important in the broader aspects of linguistics, are not of interest in this work.

3.3.3. Definition of "Sentence"

From the foregoing definitions a SENTENCE can now be defined. A SENTENCE is an ordered triple N_i, R, N_j where N_i, N_j are NAMES (SIMPLE, COMPOSITE, COMPLEX or vacuous) and where R is a PRIMARY RELATION (SIMPLE or COMPOSITE, never vacuous). If N_i, N_j do not themselves contain PRIMARY RELATIONS, then this definition of "sentence" is equivalent to Cook's definition of "clause"⁷ (12). I shall use Cook's definition of "clause," so that "sentence" is superordinate to "clause."

3.3.4. Interregnum

Let me give here an informal summary of the elements of the theory of language I am proposing. The basic elements of a language are things called symbols or alphabetical symbols. From these symbols, strings (linear sequences) of them may be formed which are called words. The set of such words is called a vocabulary. Some of these words name things and I have called these words NAMES. Some words name behaviors of things, and I have called these words RELATIONS. More complex NAMES and RELATIONS may be produced by stringing together simpler NAMES and

7. Cook states that a clause is a string of words with one and only one predicate (in my terms predicate = PRIMARY RELATION).

RELATIONS. A SENTENCE was defined as an ordered triple $N_1 R N_j$ such that R is a PRIMARY RELATION. Given the special condition that the sentence contains but one PRIMARY RELATION, a CLAUSE was defined. In the next section I will show, for English, the way in which these various linguistic elements are employed, in order to complete the theory. The various terms so far defined are summarized in Table 2.2. A decomposition of a sentence based on the concepts presented is given in Figure 2.2.

3.4. Formal Statement of a Theory of Language

In Section 3.3 I have defined and exemplified certain notions basic to the theory of language I propose. The way in which these basic notions are interrelated in language utterances must now be specified in order to produce a coherent theory of language (at least of English). I shall first consider the function of RELATIONS and then shall define successively larger aggregates of NAMES and RELATIONS culminating in the definition of SENTENCE.

3.4.1. PRIMARY RELATIONS

In Section 3.3.2, RELATIONS were categorized as primary or secondary, simple or composite, dominant or recessive. The relationships between these categories of RELATION are illustrated in Figure 2.3. Some additional terminology is needed to facilitate further discussion of PRIMARY RELATIONS. COMPOSITE PRIMARY RELATIONS consist of a series of PRIMARY RELATIONS. Let us call the right-most RELATION a MAIN RELATION and the other RELATIONS in the series ALLIED RELATIONS. Thus, a PRIMARY RELATION may be characterized as

Table 2.2 Summary of Defined Terms.

Term	Abbreviation	Definition
PRIMARY RELATION		a single word which is assigned to a behavior.
SECONDARY RELATION		a special linguistic element which has no referent in things or behaviors but serves to relate NAMES to NAMES or to relate RELATIONS to NAMES or RELATIONS.
DOMINANT PRIMARY RELATION	DP-RELATION	a PRIMARY RELATION which can serve in isolation to relate one or more NAMES.
RECESSIVE PRIMARY RELATION	RP-RELATION	a PRIMARY RELATION which is an argument of another PRIMARY RELATION.
SIMPLE RELATION		a single PRIMARY or SECONDARY RELATION.
COMPOSITE SECONDARY RELATION	CS-RELATION	a series of SECONDARY RELATIONS
COMPOSITE PRIMARY RELATION	CP-RELATION	a series of PRIMARY RELATIONS
COMPOSITE DOMINANT PRIMARY RELATION	GDP-RELATION	a series of DP-RELATION or a series of DP-RELATIONS preceded or followed by a CS-RELATION.
COMPOSITE RECESSIVE PRIMARY RELATION	CRP-RELATION	a series of RP-RELATIONS or a series of RP-RELATIONS followed by a CS-RELATION

Table 2.2 (continued)

Term	Abbreviation	Definition
SIMPLE NAME		a single NAME
COMPOSITE NAME		a series of single NAMES
COMPLEX NAME		a NAME - RELATION - NAME triple
CLAUSE		an ordered NAME - RELATION - NAME triple such that "RELATION" contains a PRIMARY RELATION and neither NAME contains a PRIMARY RELATION
SENTENCE		an ordered NAME - RELATION - NAME triple such that "RELATION" contains a PRIMARY RELATION

SENTENCE: The routine microscopic blood inspection that is a universal feature of present medical practice is generally depended on as one of the most useful indicators of the state of a person's health.

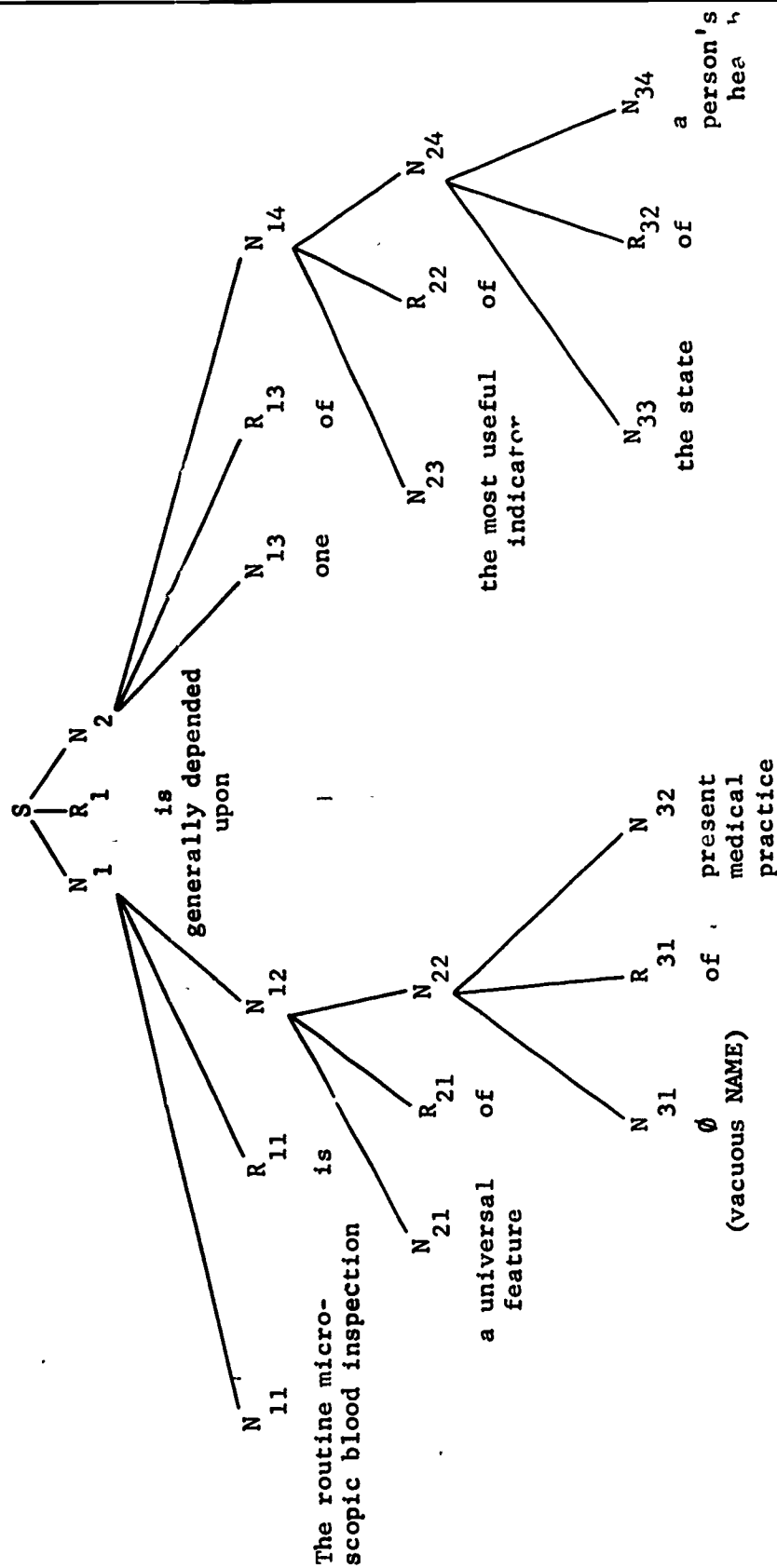


Figure 2.2 An example of a sentence analyzed in terms of NAMES and RELATIONS. COMPOSITE NAMES have not been decomposed. 24

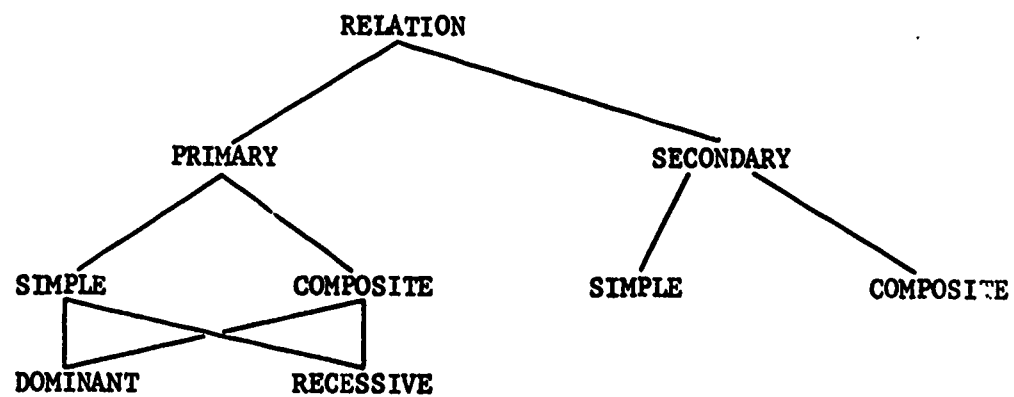


Figure 2.3 Relationships between categories of RELATION.

(ALLIED RELATION)_n MAIN RELATION

($n \geq 0$)

If $n > 0$, then the string constitutes a COMPOSITE PRIMARY RELATION.

It will be convenient to further partition ALLIED RELATIONS into three classes: AUXILIARYs, MODALs and ADJUNCTs. These three classes are each ostensibly defined as in Table 2.3. It can be seen that these three classes of ALLIED RELATION are quite similar to the traditional classes: auxiliary verbs, modal verbs and adjunct verbs.

MAIN RELATIONS may be either dominant or recessive (this distinction does not apply to ALLIED RELATIONS). The purpose of this distinction will be made clear shortly. RECESSIVE MAIN (RM) RELATIONS are of two types: PARTICIPIAL or INFINITIVAL. PARTICIPIAL RM-RELATIONS are SIMPLE RELATIONS ending in "ing," except for those mentioned in Table 2.4.⁸ INFINITIVAL RM-RELATIONS are classed as marked or unmarked. Marked INFINITIVAL RM-RELATIONS are those RM-RELATIONS which immediately follow the SECONDARY RELATION "to." All other INFINITIVAL RM-RELATIONS are ⁹unmarked.

-
8. The definition given of a PARTICIPIAL RM-RELATION must be distinguished from the traditionally accepted definition of participle. A PARTICIPIAL RM-RELATION is defined as the name of a behavior. Thus, the word "barking" in the phrase "the barking dog" is part of a COMPOSITE NAME, not a PARTICIPIAL RM-RELATION.
 9. As will be seen, this is not an adequate definition of unmarked INFINITIVAL RM-RELATIONS, for analytical purposes. No satisfactory general definition has so far been found.

Table 2.3 Classes of ALLIED RELATION and their Elements.¹⁰

AUXILIARY	MODAL	ADJUNCT
am	can	did
are	could	do
be	may	does
been	might	keep
being	must	kept
is	shall	get
had	should	got
has	will	let
have	would	
having		
was		
were		

10. It is emphasized that the elements listed may also be found to be elements of the set of MAIN RELATIONS; the three classes are mutually exclusive only within the framework of ALLIED RELATION.

Table 2.4 RELATIONS ending in "ing" which are not
PARTICIPIAL RM-RELATIONS.

bring	ring	cling	string
ding	sing	ping	swing
fling	spring	bing	wing

The complete hierarchy of PRIMARY RELATIONS is given in Figure 2.4.

3.4.2. SECONDARY RELATIONS

As stated earlier, SECONDARY RELATIONS are special linguistic devices which have no extralinguistic referents. However, SECONDARY RELATIONS play an important role in gaining an understanding of English. Therefore, it is necessary to make a number of subdivisions of SECONDARY RELATIONS. This is done in the following section. Some initial indication of the functionings of SECONDARY RELATIONS is also introduced at this time.

SECONDARY RELATIONS are initially divided into two classes: CONJUNCTIVE and ATTRIBUTIVE; these two classes are discussed, in turn, below.

3.4.2.1. The CONJUNCTIVE SECONDARY RELATIONS

CONJUNCTIVE SECONDARY RELATIONS (CSR) are categorized according to function. The categories are named COORDINATE, SUBORDINATE, ADVERBIAL, NOMINAL and ADJECTIVAL. Each of these categories is defined in the following paragraphs.

There are just five COORDINATE CSRs:

and	not	nor
or	but	

SUBORDINATE CSRs are:

if	however	although
than	therefore	thus
then	though	whether
since	yet	unless

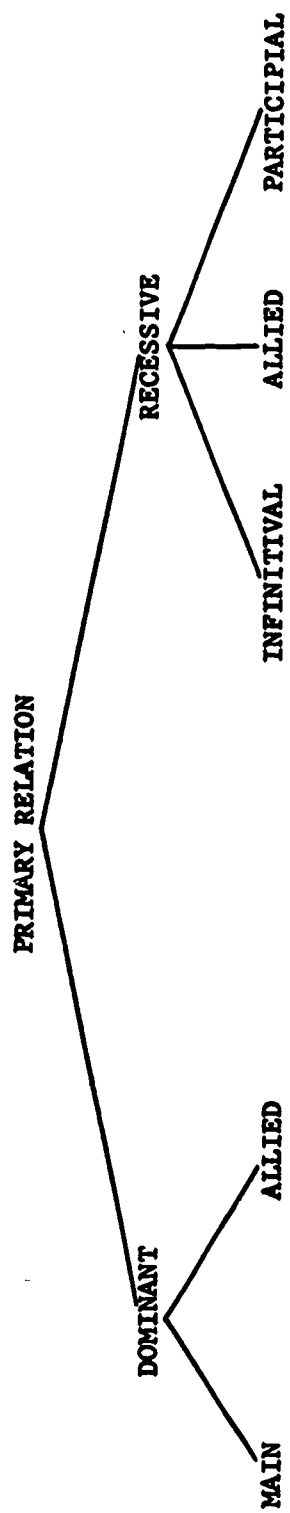


Figure 2.4 The complete hierarchy of PRIMARY RELATIONS.

Seven elements make up the set of ADVERBIAL CSRs:

where	from	to
where	by	
in	for	

The following words constitute the set of NOMINAL CSRs:

what	whoever	that
which	whomever	whom
why	whatever	of
how		

Finally, the ADJECTIVAL CSRs are:

who	when	of
whom	that	
where	whose	

Obviously these classes of CONJUNCTIVE SECONDARY RELATIONS are neither mutually exclusive among themselves nor with respect to other classes of SECONDARY RELATIONS (see below). The means of distinguishing between the classes given a particular element will be dealt with in Section 3.4.3. First let me treat ATTRIBUTIVE SECONDARY RELATIONS.

3.4.2.2. The ATTRIBUTIVE SECONDARY RELATIONS

ATTRIBUTIVE SECONDARY RELATIONS (ASR) are listed below.

of	through	above
to	down	across
in	between	outside
for	under	except
with	off	beyond
on	during	inside
at	without	instead
by	around	throughout
from	upon	despite
out	until	about
up	toward	into
over	among	below
after	within	according
before	along	behind

ASRs serve either an ADJECTIVAL or an ADVERBIAL function. ADJECTIVAL ASRs are "of" and those ASRs which follow a DOMINANT RELATION and precede a SIMPLE or COMPOSITE NAME. All other ASRs are ADVERBIAL.

The hierarchy of SECONDARY RELATIONS is shown in Figure 2.5.

3.4.3. NAMES

The concept NAME was defined in Section 3.3. It will be convenient to expand upon this concept both to provide additional useful terminology and to make clear certain functional distinctions between SECONDARY RELATIONS as alluded to above.

3.4.3.1. The Notion of PHRASE

The term PHRASE will be applied as follows. A SIMPLE or COMPOSITE NAME will be called a NOMINAL PHRASE. Thus, any element in either of the first two columns of Table 2.1, for example, amounts to a NOMINAL PHRASE.

If either a SIMPLE or COMPOSITE NAME is immediately preceded by (i.e., is an argument of) an ATTRIBUTIVE SECONDARY RELATION, and the RELATION/NAME NOMINAL is called a SECONDARY PHRASE.

A COMPOSITE PRIMARY RELATION is called a PRIMARY PHRASE.

A hierarchical arrangement of PHRASE types is shown in Figure 2.6. Examples of them are given in Table 2.5.

3.4.3.2. The Notion of CLAUSE

A COMPLEX NAME containing one and only one PRIMARY RELATION is a PRIMARY COMPLEX NAME or CLAUSE. There are four basic types of CLAUSE: PRINCIPAL, NOMINAL, ADVERBIAL and ADJECTIVAL, related as illustrated in Figure 2.7.

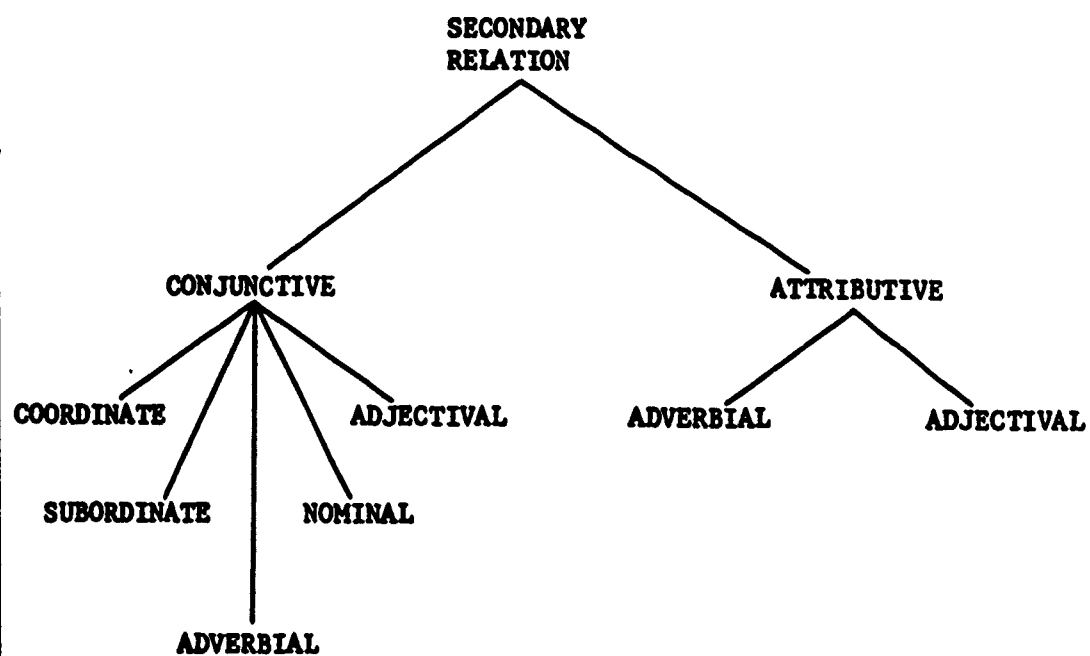


Figure 2.5 Hierarchical partitioning of SECONDARY RELATION.

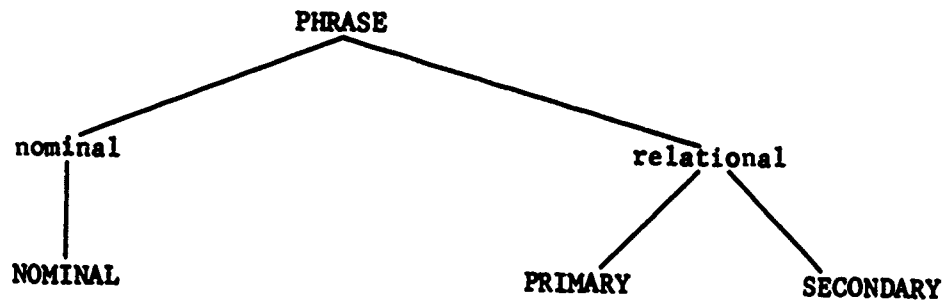


Figure 2.6 Partitioning of the class of PHRASEs.

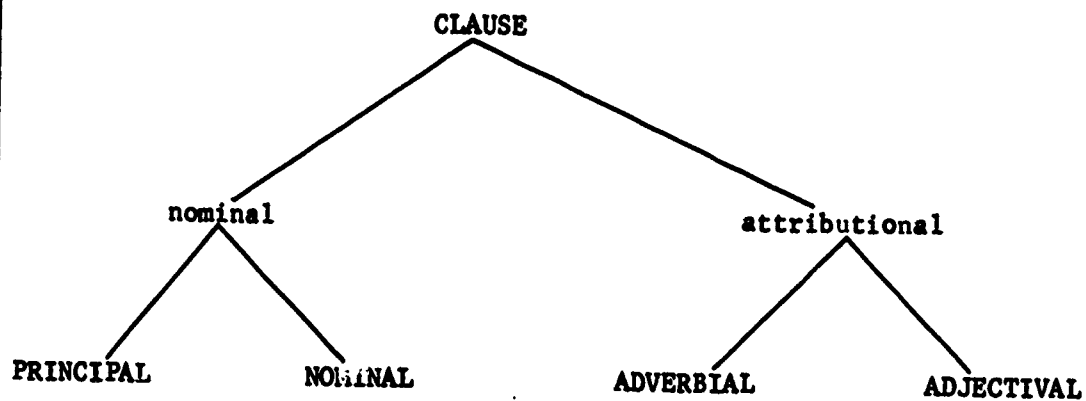


Figure 2.7 Partitioning of the class of CLAUSES.

Table 2.5 Examples of NAMES, PHRASES and CLAUSES

Term	Illustrative Example
COMPLEX NAME COMPOSITE NAME SIMPLE NAME	the book on the table the thick blue book book
NOMINAL PHRASE PRIMARY PHRASE SECONDARY PHRASE	the thick blue book has been gone in the thick blue book
ADJECTIVAL CLAUSE ADVERBIAL CLAUSE NOMINAL CLAUSE PRINCIPAL CLAUSE	the girl <u>sitting in the chair</u> the birds flew <u>where the weather was warm</u> <u>flying planes</u> can be dangerous <u>the musician became proficient</u> by practicing every day.

3.4.4. SENTENCE

Thus far, I have written at length about NAMES and RELATIONS, and I have given some indication as to the way¹¹ these basic language elements are combined to form more complex units. But it still remains to consider the ways¹¹ in which these units are combined to produce the fundamental element of written English: the SENTENCE.¹²

Traditionally, a sentence has been thought of as a string of words expressing a complete thought (13). More recently, a sentence has been viewed as a string of words containing at least one predicate (14), or as any string of finite length composed of symbols from an alphabet (15). In this research, the simplest form of sentence has been the CLAUSE. If one considers that a RELATION demands a certain number of NAMES as arguments and that these arguments may themselves contain other RELATIONS and their arguments, it is easy to see that English is a relational language after the notions of Rothstein (16).

I would like to think of SENTENCE as a NAME. Of course, SENTENCES may be of varying complexities, ranging from just a single PRIMARY RELATION to strings containing many NAMES and RELATIONS. But the problem is to specify the way in which relatively simple language units may be combined to yield SENTENCES. From my point of view, it will suffice to show that,

11. But not why they are so combined.

12. Communication through written English does employ yet higher aggregations of linguistic units, but I shall not attempt to incorporate them into my theory of language at this time (see, however, Strong (17)).

given a string of English words, rules can be devised within the framework of the theory of language I have so far presented that permit the identification of the linguistic units specified within the theory and that, as a consequence, permit one to say whether the given string is or is not an English sentence.

Since, within the theoretical framework I have postulated, all types of relations have been identified and defined, I shall define SENTENCE quasi-ostensively through specification of the argument(s) each type of RELATION may take. The list of RELATIONS to be dealt with is given in Table 2.5, together with the range of arguments each may take. In order to express the RELATION/argument in a simple way I shall adapt Griswald's prefix notation (18) for my purposes. A few words are in order concerning this notation.

In prefix notation, the operator (relation) occurs first, followed by the arguments which it relates. Thus, the expression $A + B$ would become $+AB$. The number of arguments associated with an operator is specified by a numeral subscripted to the operator. Thus, the previous expression becomes $+_2AB$. To avoid ambiguity in the application of operators, a precedence is assigned to them (i.e., they are given a priority for first application).

For my purposes, 15 operators are employed, corresponding to the RELATIONS already defined. Table 2.7 lists the operator symbols which I shall use, together with their precedence and the RELATION to which each symbol corresponds. Each sentence type defined within the theory I propose is given in Table 2.8 together with an example English

Table 2.6 Definition of Prefix-Notation Symbols as Used in the Specification of SENTENCES.

SYMBOL	NUMBER OF ARGUMENTS (n)	PRECEDENCE	USE
a	$0 \leq n \leq 3$	2	MAIN PRIMARY RELATION
b	$0 \leq n \leq 3$	2	AUX and VRB or AUX and MOD
c	$0 \leq n \leq 3$	2	AJN and VRB
d	$0 \leq n < 3$	2	AJN
e	$0 \leq n \leq 3$	2	PTC
f	$0 \leq n \leq 3$	2	INF
g	$0 \leq n \leq 3$	2	AUX and INF or MOD and INF
h	$n = 2$	1	ADJECTIVAL ATTRIBUTIVE SECONDARY (AJAS) RELATION
i	$n = 2$	3	ADVERBIAL ATTRIBUTIVE SECONDARY (AVAS) RELATION
j	$n \geq 2$	4	CCP
k	$n \geq 2$	1	CCN
l	$n = 2$	4	SCN
m	$n = 2$	3	ADJECTIVAL CONJUNCTIVE SECONDARY (AJCS) RELATION
n	$n = 2$	3	ADVERBIAL CONJUNCTIVE SECONDARY (ADCS) RELATION
ø	$n = 0$	1	NOMINAL CONJUNCTIVE SECONDARY (NCS) RELATION
<hr/>			
'VERB'		-	Any from of specific verb
'word'		-	A specific lexical entry
N		-	An upper case letter represents a NAME

Table 2.7 Specification of SENTENCES.

SENT- ENCE TYPE	RELA- TION TYPE	ARGU- MENTS	NOTATION FOR SENTENCE	ILLUSTRATIVE EXAMPLE
1	a	0	a_0	Go. a
2	b	1	bA_1	The boy is going. A b
3	a	2	aA_1	Throw the ball. a A
4	a	2	aAB_2	The boy threw the ball. A a ₂ B
5	a	2	aAB_2	Give the boy the letter. a ₂ A B
6	a	2	aAB_2	Flying planes are dangerous. ¹³ A a ₂ B

¹³ The example has been inserted to show that what is traditionally defined to be a participle (contrary to the definition of PTC in this work) is treated as part of a COMPOSITE NAME.

Table 2.7 (continued)

SENT- ENCE TYPE	RELA- TION TYPE	ARGU- MENTS	NOTATION FOR SENTENCE	ILLUSTRATIVE EXAMPLE
7	a	3	$a_3 ABC$	The boy gave me the letter. A a_3 B C
8	a	3	$a_3 ABC$	They named him president. A a_3 B C
9	d	1	$d_1 g B$	Let him go. d_1 B g_1
10	d	2	$d_2 Ag B$	The boy let him go. A d_2 B g_1
11	d	1	$d_1 g A$	Let go of him. d_1 g_1 A
12	d_2	2	$d_2 Ai a BC$	The teachers kept the children working in the classroom. A d_2 B a_1 i_2 C
13	d	1	$d_1 i a BC$	Keep the children working in the classroom. d_1 B a_1 i_2 C

Table 2.7 (continued)

SENT- ENCE TYPE	RELA- TION TYPE	ARGU- MENTS	NOTATION FOR SENTENCE	ILLUSTRATIVE EXAMPLE
14	d	1	$d \begin{smallmatrix} a \\ 1 \end{smallmatrix} 0$	Keep working. $\underbrace{d}_{1} \begin{smallmatrix} a \\ 0 \end{smallmatrix}$
15	e	1	$a \begin{smallmatrix} h \\ 2 \end{smallmatrix} \begin{smallmatrix} Ae \\ 2 \end{smallmatrix} \begin{smallmatrix} BC \\ 1 \end{smallmatrix}$	The process of making bricks requires intense heat. $\underbrace{A}_{2} \begin{smallmatrix} h \\ 2 \end{smallmatrix} \begin{smallmatrix} e \\ 1 \end{smallmatrix} \begin{smallmatrix} B \\ a \end{smallmatrix} \begin{smallmatrix} 2 \\ C \end{smallmatrix}$
16	e	1	$a \begin{smallmatrix} e \\ 2 \end{smallmatrix} \begin{smallmatrix} AB \\ 1 \end{smallmatrix}$	Flying planes is dangerous. $e_1 \begin{smallmatrix} A \\ a \end{smallmatrix} \begin{smallmatrix} 2 \\ B \end{smallmatrix}$
17	e	1	$b \begin{smallmatrix} Ae \\ 2 \end{smallmatrix} \begin{smallmatrix} B \\ 1 \end{smallmatrix}$	The women were tired of making pastries. $\underbrace{A}_{2} \begin{smallmatrix} b \\ 2 \end{smallmatrix} \begin{smallmatrix} \phi \\ e \end{smallmatrix} \begin{smallmatrix} 1 \\ B \end{smallmatrix}$
18	f	2	$a \begin{smallmatrix} Af \\ 2 \end{smallmatrix} \begin{smallmatrix} BC \\ 2 \end{smallmatrix}$	He wanted him to lift weights. $A \begin{smallmatrix} a \\ 2 \end{smallmatrix} \begin{smallmatrix} B \\ 1 \end{smallmatrix} \begin{smallmatrix} f \\ 1 \end{smallmatrix} \begin{smallmatrix} C \end{smallmatrix}$
19	f	1	$a \begin{smallmatrix} Af \\ 2 \end{smallmatrix} \begin{smallmatrix} C \\ 1 \end{smallmatrix}$	He wanted to lift weights. $A \begin{smallmatrix} a \\ 2 \end{smallmatrix} \begin{smallmatrix} f \\ 1 \end{smallmatrix} \begin{smallmatrix} C \end{smallmatrix}$
20	f	0	$a \begin{smallmatrix} Af \\ 2 \end{smallmatrix} \begin{smallmatrix} 0 \end{smallmatrix}$	He wanted to leave. $A \begin{smallmatrix} a \\ 2 \end{smallmatrix} \begin{smallmatrix} f \\ 0 \end{smallmatrix}$

Table 2.7 (continued)

SENT- ENCE TYPE	RELA- TION TYPE	ARGU- MENTS	NOTATION FOR SENTENCE	ILLUSTRATIVE EXAMPLE
21	f	2	d Af BC 2 2	He let him lift weights. A d ₂ B f ₂ C
22	f	1	d Af B 2 1	He let him go. A d ₂ B f ₁
23	f	0	d Af 2 0	He let go. A d ₂ f ₀
24	g	0	a An g B 2 2 0	He wanted to be known as a great artist. A a ₂ g ₀ n ₂ B
25	h	2	a h ABC 2 2	The boy in the red shirt hit the target. A h ₂ B a ₂ C
26	h	2	b Ah BC 2 2	The student was taught the history of famous artists. A b ₂ B h ₂ C
27	i	2	i b AC 2 1	The papers were thrown into the waste basket. A b ₁ i ₂ C

Table 2.7 (continued)

SENT- ENCE TYPE	RELA- TION TYPE	ARGU- MENTS	NOTATION FOR SENTENCE	ILLUSTRATIVE EXAMPLE
28	i	2	i i a ABC 2 2 1	<p>The ball rolled into the street under the car.</p> <p>A a₁ i₂ B i₁ C</p>
29	j	2	a j ABC 2 2	<p>Bob and Carol built a house.</p> <p>A j₂ B a₂ C</p>
30	j	2	i j a a AB 2 2 1 1	<p>The little boy ran and skipped into the house.</p> <p>A a₁ j₂ a₁ i₂ B</p>
31	j	3	a A j BCD 2 3	<p>The children liked cats, dogs and gerbils.</p> <p>A a₂ B C j₃ D</p>
32	j	2	i i a ABC 2 2 1	<p>The girls ran out of the house and into the street.¹⁴</p> <p>A a₁ i₂ B j₂ i₁ C</p>

14. When a CCP joins SECONDARY PHRASES initiated by an AVAS-RELATION, the C P is omitted from the notation.

Table 2.7 (continued)

SENT- ENCE TYPE	RELA- TION TYPE	ARGU- MENTS	NOTATION FOR SENTENCE	ILLUSTRATIVE EXAMPLE
33 ¹⁵	k	2	$k \begin{smallmatrix} i & a \\ 2 & 2 \end{smallmatrix} \begin{smallmatrix} 1 & 1 \end{smallmatrix} AC(A) \begin{smallmatrix} i & a \\ 2 & 1 \end{smallmatrix} B$	<p>The dogs $\underbrace{\text{raced down the hill}}_C$ and $\underbrace{\text{jumped into the water.}}_D$</p> <p>$\underbrace{A}_{a \begin{smallmatrix} 1 \end{smallmatrix}}$ $\underbrace{k \begin{smallmatrix} 2 \end{smallmatrix} a \begin{smallmatrix} 1 \end{smallmatrix} i \begin{smallmatrix} 2 \end{smallmatrix}}_D$</p> <p>The little dog $\underbrace{\text{laughed and the dish ran away with}}_B$</p> <p>$\underbrace{A}_{a \begin{smallmatrix} 1 \end{smallmatrix}}$ $\underbrace{k \begin{smallmatrix} 2 \end{smallmatrix} B \begin{smallmatrix} a \begin{smallmatrix} 1 \end{smallmatrix} i \begin{smallmatrix} 2 \end{smallmatrix}}_D$</p> <p>the spoon.</p> <p>\underbrace{C}</p>
35 ¹⁶	1	2	$1 \begin{smallmatrix} a & Ab & B \\ 2 & 1 & 1 \end{smallmatrix}$	<p>The experiment $\underbrace{\text{failed because the equipment was faulty.}}_B$</p> <p>$\underbrace{A}_{a \begin{smallmatrix} 1 \end{smallmatrix}}$ $\underbrace{1 \begin{smallmatrix} 2 \end{smallmatrix} B \begin{smallmatrix} b \begin{smallmatrix} 1 \end{smallmatrix} \end{smallmatrix}}_B$</p>
36	1	2	$1 \begin{smallmatrix} a & Ab & B \\ 2 & 1 & 1 \end{smallmatrix}$	<p>Because the equipment $\underbrace{\text{was faulty the experiment failed.}}_B$</p> <p>$\underbrace{1 \begin{smallmatrix} 2 \end{smallmatrix} B \begin{smallmatrix} b \begin{smallmatrix} 1 \end{smallmatrix} A \begin{smallmatrix} a \begin{smallmatrix} 1 \end{smallmatrix} \end{smallmatrix}}_B$</p>

15. In this SENTENCE the case is illustrated where one argument must be repeated in the formalization. the second occurrence of the argument is represented by enclosing the symbol in parentheses -- "(A)".

16. SENTENCES 35 and 36 illustrate that the operation representing SCN RELATION is followed first by the PRINCIPAL CLAUSE and then by the CLAUSE initiated by the SCN.

Table 2.7 (continued)

SENT- ENCE TYPE	RELA- TION TYPE	ARGU- MENTS	NOTATION FOR SENTENCE	ILLUSTRATIVE EXAMPLE
37 ¹⁷	m	2	$a_2(m_2)A_1b_2BCD$	The girl who was sitting in the first row won the prize $\begin{array}{ccccccc} A & B & b_1 & i_1 & C & a_2 & D \\ & (m_2) & & & & & \end{array}$
38	n	2	$n_2a_1Aa_2AB$	They stopped when they saw the smoke. $\begin{array}{ccccccc} A & a_1 & n_2 & A & a_2 & B & \\ & & & & & & \end{array}$
39	ϕ	-	a_1Ab_2BC	He knew that she was telling the truth. $\begin{array}{ccccccc} A & a_1 & \phi & B & B_2 & C & \\ & & & & & & \end{array}$

17. The ADJECTIVE COORDINATE SECONDARY RELATION " (m_2) " is placed in parentheses because REL_1 serves both as a NAME and a RELATION.

sentence for illustration. The following discussion should clarify the definition of SENTENCES in these terms. Consider the SENTENCE

The equipment in this room is used for testing compounds.

$\underbrace{\text{A}} \quad \underbrace{\text{h}_2} \quad \underbrace{\text{B}} \quad \underbrace{\text{b}_1} \quad \underbrace{\text{n}_2} \quad \underbrace{\text{e}_1} \quad \text{C}$

Each SIMPLE or COMPOSITE NAME is represented as a unique upper-case letter.

The SENTENCE contains four RELATIONS: "h" is an AJAS-RELATION with 2 arguments, "b" is a DP-RELATION with 1 argument, "n" is an ADCS-RELATION with 2 arguments and "e" is a RP-RELATION. To formalize the SENTENCE, begin with the highest precedence operator, i.e., the AJAS-RELATION "in," which relates "A" and "B" to form the COMPLEX NAME "h₂AB." The operator applied next is the PRIMARY RELATION operator. Since there are two, simply apply the operators in a left-to-right manner. "The equipment is use" becomes the COMPLEX NAME "b₁h₂AB" and "testing compounds" becomes the COMPLEX NAME "e₁C." Finally, apply the ADCS-RELATION operator to obtain the COMPLEX NAME "n₂b₁h₂ABe₁C."

References

1. A. Kaplan, The Conduct of Inquiry, Chandler Publishing Company, San Francisco, California, 1964, p. 428.
2. ibid., p. 308.
3. ibid., p. 308.
4. D. E. Knuth, The Art of Computer Programming, Addison-Wesley Publishing Company, Massachusetts, 1968, p. 17.
5. R. Simmons, J. Burger and R. Schwarcz, "A Computational Model of Verbal Understanding," Proceedings of the Fall Joint Computer Conference 33 (1), 441-456 (1968).
6. J. Rothstein, "Relation Languages," Report to National Science Foundation Office of Science Information Service, Section 5.8, The Ohio State University, Columbus, Ohio., 1970, pp. 5.44-5.46.
7. J. Rothstein, "A Class of Formal Languages for Possible Biological Application," First European Biophysics Congress, 1971.
8. Simmons, et al., op. cit.
9. C. A. Montgomery, "Linguistics and Information Science," American Society for Information Science 23(3), 195-219 (1972).
10. C. J. Fillmore, "The Case for Case," in Universals in Linguistic Theory, Bach, et al. (ed.), Holt, Rinehart and Winston, Inc., New York, 1968.
11. W. L. Chafe, Meaning and the Structure of Language, The University of Chicago Press, Chicago, Illinois, 1970.
12. W. A. Cook, S.J., Introduction to Tagmemic Analysis, Holt, Rinehart, and Winston, Inc., New York, 1969, p. 65.
13. A. Bain, A Higher English Grammar, London, 1879, in C. Fries, The Structure of English, Harcourt, Brace & World, Inc., New York, 1952.
14. Cook, op. cit., p. 65.
15. S. Ginsberg, "Lectures on Context-Free Languages," in M. A. Arbib (ed.), Algebraic Theory of Machines, Languages, and Semigroups, Academic Press, New York, 1968, pp. 313-327.

16. ibid., Rothstein, op. cit., 1971.
17. S. M. Strong, "An Algorithm for Generating Structural Surrogates of English Text," Masters Thesis, The Ohio State University, March, 1973.
18. R. E. Griswald, SNOBOL4 -- Structure and Implementation, Presented at SHARE XXXVII, New York City, August 12, 1971.

CHAPTER III. IDENTIFICATION OF SIMPLE NAMES AND RELATIONS

People studying a foreign language always worry about vocabulary. They will stare incredulously at a teacher who tells them that vocabulary is not the most important part of learning a foreign language. Occasionally a student may understand why this is true, but only after he has laboriously looked up all the words in a passage and still finds that he can make no sense out of the assembled words.

Ann Eljenholm Nichols, English Syntax

1. Introduction

In the preceding chapter a theoretical framework was developed within which language was viewed as a relational system. Thus, two classes of linguistic element were demanded: NAMES and RELATIONS. In this and succeeding chapters I shall attempt to show the usefulness of this theory as a guide in developing algorithms which identify and label various components of an English text. In this chapter, an algorithm is described which classifies words into several classes of NAME or RELATION according to the structural properties of the (usually) sentence examined. Before describing the algorithm, however, a review of some related work is presented in order to provide some context within which to view the present research.

2. Overview of Related Work in Syntactic Analysis

Syntactic analysis has been studied by many groups of researchers. In this section are reviewed those studies in syntactic analysis most pertinent to the present work. The section is divided into two parts.

The first part contains descriptions of four computer programs for syntactic analysis which, although rather large and complex, are well-grounded theoretically. In the second part are described six programs which bear especially close similarity to the program developed in this research.

Only superficial comparisons among these programs can be made. Each program has been developed to identify a unique set of grammatical classes; each has been designed based on a unique set of objectives; each has been implemented using a different programming language, and on a different computer. Furthermore, the accuracy of the results produced by these programs is not always available. Thus, these brief summaries are written to give the reader an idea of what has been accomplished and of the obstacles which remain to the improvement of algorithms for syntactic analysis.

2.1. Review of Syntactic Analysis Procedures--Part I

The four studies presented in this section give the reader a general overview of much of the research into syntactic analysis carried out to date. The procedures developed in each of the four projects have the common attribute of being quite large and complex. In each procedure a large lexicon (dictionary) is employed, and it will be seen that the procedures were developed without any attempt at a distinction between structural properties of a language and the intensional properties of the language. The nature of the procedures also necessitates that the lexicon be exhaustive. And in general, these procedures produce all possible analyses of a sentence from which one or more preferred ones

may be selected, of course manually.

All of the programs assign words to syntactic classes, but because of the nature of the grammar upon which the programs are based, word-class assignment is inextricably linked to phrase-group and clause-group assignment. For this reason, some of these programs will be mentioned again briefly in the succeeding chapter in which clause and phrase analysis is discussed. The investigations discussed in this section are those headed by Kuno and Oettinger, Zwicky, Sager, and Winograd.

2.1.1. The Multiple-Path Syntactic Analyzer

A well-known program for syntactic analysis is the Multiple-Path Syntactic Analyzer developed by Kuno and Oettinger (1). This system is based on a context-free grammar of 3400 rules and a top-to-bottom (top-down) analytical procedure employing a pushdown store. The dictionary used by this procedure gives a highly refined division of syntactic classes. For example, "are" belongs to three syntactic classes: one when used as an intransitive verb, one when used as a finite copula and one when used as an auxiliary to another verb. Each possible analysis of the sentence is explored in a left-to-right manner and verified or invalidated by the context-free grammar. The production of multiple analyses is useful for research purposes, but it is a decided disadvantage in practical application. Processing time is not directly dependent upon the length of a sentence but depends primarily on the number of possible surface structures which the sentence can generate (2). As an example, one 35-word sentence was reported to

require 12 minutes for analysis (3).

2.1.2. Syntactic Analysis for Transformational Grammar

Zwicky (4) has included syntactic analysis in an attempt to implement a transformational grammar for sentences. The first phase of the program is a phrase-structure grammar, which, at present, handles a subset (32 rules out of a possible 134 rules have been implemented) of English. The initial step in Zwicky's procedure consists of a dictionary look-up of each word in a sentence. The dictionary contains all possible syntactic classes for a word, along with attributes such as tense, transitivity, and number. This lexical entry also represents an entry in terms of more abstract elements, for example, "none" is defined as "neg any." After all possible syntactic classes for a word in a sentence have been retrieved, all possible analyses of a sentence are examined. The sentence

Can the airplane fly.

has 15 possible surface structures, since "can" has 5 entries, airplane has 1 entry and fly has 3 entries. In the next step a context-free grammar is applied to each of these analyses, and some structures are eliminated. Transformational rules are applied next in an effort to eliminate the spurious surface structures. The grammar is capable of analyzing sentences with more than one embedded clause. No discussion of the lexicon and of the rules used in both the context-free grammar and the transformational grammar is given. Although the process is

only partially implemented,¹⁸ Zwicky concludes that highly efficient routines are needed to obtain the correct surface structures (6).

2.1.3. Syntactic Analysis Based on String Analysis

In a project headed by Naomi Sager a string analysis grammar (7) is the basis of a syntactic analysis program. In this analysis, a sentence is viewed as a set of elementary strings (clauses). Modifiers and prepositional phrases are defined as adjuncts. Thus, the sentence

Cars without brakes cause accidents.

is described by the elementary string - N tV N (i.e., noun, tensed verb, noun) and an adjunction class P N (preposition, noun). Rules are developed for the analysis of relative clauses and of clauses joined by coordinate conjunctions. The program utilizes a dictionary which identifies all possible syntactic classes of each word. In the first step of the program, the possible syntactic classes of each word in the sentence are retrieved. The assignments made during this step are:

Cars without brakes cause accidents.
N P N/tV N/tV/V N

In the next phase of the program, the syntactic categories are examined in a left-to-right manner, while the grammar rules for elementary strings and adjuncts are matched to the possible analyses of the sentence.¹⁹ If a grammar rule fails to match up correctly with the

18. The lexical look-up and context-free parsing steps have not been programmed (5).

sentence, the program backs up to the initial word of an adjunct or string and attempts to apply a different rule. Sager estimates that an adequate grammar for English could be accomplished with about 150 rules, plus another 150 restrictions. These restrictions analyze elements of the elementary string and check for such things as subject-verb agreement and well-formedness. No discussion is given of the timings, storage, the accuracy of the program or of the rules which have been developed (8).

2.1.4. Syntactic Analysis for the Simulation of Natural Language Processes

Winograd has recently published a system design to explore how humans process language (9). The system includes procedures for both syntactic and semantic analysis. The syntactic analysis procedures are based on a systemic theory of grammar (10, 11). Although the theories of this grammar have not been stated in a unified way, the emphasis of this grammar is on the "informational units" (12) of a language. Winograd interprets these "informational units" as amounting to clauses and phrases.

In his procedures for syntactic analysis, Winograd has defined 18 word classes, 4 group types (noun, verb, adjective and preposition) and 2 clause types. The 18 word classes are a finer division of the traditional word classes. The two types of clauses are denoted as major and secondary. A major clause is either imperative, declarative or a question. Associated with each of these units is a set of attributes. Every unit belonging to a word class, group or clause is assigned a subset of these attributes. For example, a verb (e.g., "began") may have the attributes "past," "infinitive;" a preposition group (e.g., "in the kitchen") may have the attribute "locational object"; and a clause may have attributes

such as "transitive," "passive" and "causality."

The components of the syntactic analysis process are a dictionary, a context-free grammar and a push-down list (PDL). The dictionary contains the vocabulary of the language, along with each entry's syntactic class and attributes. The program which implements the context-free grammar contains several functions which give the grammar context-sensitive aspects. As an example, one function checks for agreement between subject and predicate. A bottom-up parser is used to apply the context-free grammar.

This process appears to represent a viable system for the context in which it is used. While the universe of discourse for the system is small, Winograd has succeeded in developing a system which has the ability to synthesize sentences and which serves as an interesting tool in the investigation of question-answering systems (13).

2.2. Review of Syntactic Analysis Procedures--Part II

The investigations presented in this section require neither exhaustive dictionaries nor particularly complex procedures. Most of these programs have been based on ad hoc rules rather than on a well-developed theory of language. The studies presented here are those headed by Clark and Wall, Klein and Simmons, Stolz, et al., Thorne, et al., Resnikoff and Dolby, and Woods.

2.2.1. The Economical Parser

In this study a program was written which assigns words to grammatical classes, identifies phrase types and marks clause boundaries (14). The program first performs a dictionary look-up. The dictionary consists of

about 1000 entries. The entries include function words, inflectional endings, and a list of words which are exceptions to regular inflection (i.e., "thing" is not a verb even though it ends in "ing"). Words which are not found in the dictionary are assigned an ambiguous noun/verb category. In a second pass, phrase boundaries are tentatively identified. In the third pass clause boundaries are identified and clauses are tested for well-formedness. If a clause does not contain a verb, noun phrases are examined in a left-to-right manner and the first word which has been assigned to the noun/verb class is identified as a verb. Nine types of phrases and eight kinds of subordinate and relative clauses are identified. The algorithm was applied to abstracts of technical material and is reported to attain 91% accuracy in the identification of grammatical classes and 91% accuracy in the identification of phrases. The processing time is given as 24 words per minute. The algorithm was written in COBOL and executed on the IBM 7094 (15).

2.2.2. The Computational Grammar Coder

Klein and Simmons have implemented a Computational Grammar Coder (CGC) which assigns words in English text to the appropriate grammatical class. The CGC is the initial phase of a syntactic analyzer which is part of a question answering system. The CGC contains two types of dictionaries: 1) a function-word dictionary containing about 400 words, and 2) two dictionaries containing those nouns, verbs, and adjectives which are exceptions in various suffix texts. For example, while most words ending in "ing" are verbs, the word swing may be a noun. These

dictionaries contained about 1,500 words.

The algorithm of the CGC begins by putting each word through a series of independent tests. These tests include a function word test, a capitalization test, a numeral test, and a series of suffix tests. Each test may result in the assignment of a set of codes to the word. If a particular test yields no information about a word, the system assumes that the classes noun, adjective, verb are possible. A final set of codes is obtained by taking the intersection of the set of codes assigned in each test.

After each word in a sentence has been identified in this manner, the context-frame test is made. This test sequentially processes strings of ambiguously coded words which are bounded by uniquely coded words. Every possible combination of codes of an ambiguously coded string is checked against a context-triad-frame table which contains permissible combinations of codes in such strings. When one of the sequences of codes of such a string is found in the context triad frame table, then this unambiguous sequence of codes is assigned. An example will clarify this process. Consider the sequence of codes:

	adjective	noun	
article	verb	adjective	verb

The first word of the string is an article, the second could be either an adjective or verb, the third either a noun or an adjective, the last is a verb. The following sequences of codes could be assigned to the ambiguously coded words:

adjective	noun
adjective	adjective
verb	noun
verb	adjective

The context-triad-frame table contains the following possible codes for two words which occur between an article and a verb:

adjective	noun
noun	adverb
noun	noun

The only sequence of codes common to the test string and to the table is ADJ - NON; thus, these are the codes which would be assigned.

This phase of the CGC is limited by two factors: 1) a string containing more than three ambiguously coded words cannot be handled and 2) out of 2,700 potential table entries, only 500 entries are included in the context-triad-frame table. The table entries were empirically derived by analyzing a chapter from the Encyclopedia Americana. The 500 entries included in the table accounted for 90% of the sequences of codes which were observed to occur in the test data (16).

The CGC, which was written in JOVIAL for an IBM 7090 uses just under 14,000 computer words. In tests using scientific text, the system correctly identified approximately 90%¹⁹ of the words (17).

19. This figure refers to the correct identification of one of 30 classes: adjective, adverb, noun, verb, the verb to be, two classes of auxiliary verbs, articles, four classes of conjunctions, four classes of prepositions, and nine kinds of pronouns, punctuation, and verbs identified by inflectional endings.

2.2.3. The WISSYN System

The WISSYN system, designed to make grammatical class assignments to words in English text, was developed by Stolz, Tannenbaum and Carstensen (18). WISSYN contains dictionaries which are similar to those used in the CGC (Section 2.2.2.). The function-word dictionary contains about 300 words. If a word in a sentence is not found in the dictionary of function words, it is checked against a dictionary of suffixes. The word must also be checked against a list of about 60 words which are exceptions to the suffix tests.

A third phase attempts to resolve the ambiguity of certain function words. For example the word "that" may be used in the following ways:

that dog jumped

the dog that jumped

In this phase, a set of frames similar to those implemented by Klein and Simmons is used to resolve residual ambiguity.

A final phase of WISSYN uses the statistical frequencies of structural patterns of English sentences to assign grammatical classes. For this phase, probabilities of individual word classes occurring in a particular structural pattern were calculated by manually analyzing English text. The operation of this phase can be easily understood by considering an illustration. Given a sentence whose elements have been identified as:

T D N X₁ P D N P X₂ X₃ T

where T is a terminal marker, D is a determiner, N is a noun, P is a

pronoun, and X_i is the i th unidentified word ($i \geq 1$). When an unidentified word is encountered, the probabilities of all strings consisting of four or fewer words surrounding the unidentified word are considered. For the above example, the following strings would be considered:

T D N X_1 P

D N X_1 P D

N X_1 P D N

X_1 P D N P

T D N X_1

D N X_1

N X_1

X_1 P D N

X_1 P D

X_1 P

D N X_1 P D

Since the longest strings provide the most reliable probabilities, only the three longest strings are actually used. Thus, the probability of X_1 being a noun, verb, adjective, or adverb would be calculated by using the following statistics:

CONDITIONAL PROBABILITIES

Predictor	NOUN	VERB	ADJ	ADV
T D N X	.046	.819	.013	.122
X P D N	.438	.359	.068	.135
D N X P N	.017	.591	.078	.314
JOINT PRODUCT	.00034	.17377	.00007	.00517

The element X_1 would be designated as a verb, since this was the most probable case. The probability tables used include only the 150 most frequent predictor configurations. When a particular configuration is not found in the table, then the table is searched for the next longest string.

In a test of literary, scientific, and newspaper articles, an accuracy as high as 93% was attained.²⁰ WISSYN has been implemented on the CDC 1604 in CDC FORTRAN 60. The program, which is contained in approximately 6500 (48-bit) words of storage, can process about 2500 words per minute (19).

2.2.4. Syntactic Analysis Based on a Regular Grammar

Thorne, Bratley, and Dewar have written a syntactic analyzer as part of a system which assigns the deep structure and surface structure to English sentences (20). In this syntactic analyzer, 5 types of sentences, 6 types of clauses, and several other syntactic categories including, gerund subjects, active verbs, modifiers and indirect objects are identified.

The analyzer employs a dictionary of fewer than 200 words. It contains function words (referred to as closed-class words), verbs, suffixes and exceptions to these suffixes. The analyzer is based upon a form of transformational grammar. The operation of the system is as follows. First, all of the function words in a sentence are identified.

20. Words were classified according to the 14 categories: nouns, verbs, adjectives, adverbs, pronouns, determiners, linking verbs, auxiliaries, intensifiers, prepositions, relative pronouns, subordinators, connectors, negatives, preverbs, and exclamations.

Based on these function words and on the information provided by the grammar as it relates to the sentence, a set of predictions is made. These predictions are tested, and based upon the results of these tests, new predictions are made for successive parts of the sentence. All of the possible predictions are tested, and every prediction which holds produces a distinct analysis for the sentence.²¹

In some sentences none of the predictions which are made are satisfied. This happens when the sentence is improperly constructed or because the grammar is not complete. In this case, a program produces a message indicating that the analyzer has failed for this sentence.

This system was implemented on the KDF-9 computer. No figures were given for the amount of storage required by the program nor for the accuracy of the results which were obtained. While no average processing times were cited for the operation of the program, several examples of sentences which had been analyzed were given, along with the time needed to process them. The example which had the longest processing time (2.285 sec) contained 9 words, while the example which had the shortest processing time (0.427 sec) contained 1 word (21).

2.2.5. A Limited Program for Syntactic Analysis

Resnikoff and Dolby have written a program for grammatical class assignment which consists of fewer than 100 COMIT instructions (22).

21. Although these authors state that the grammar employed is a context-free grammar (CFG), these predictions effectively transform it into a context-sensitive grammar (CSG). This CSG appears to be similar to that developed by Kuno and Oettinger.

This program utilizes a dictionary of 200 function words and 200 affixes. At the time their paper was published, they intended to expand the dictionary to about 1000 words. In preliminary tests on texts which include parts of Ulysses by James Joyce and a New York Times editorial, the results were reported as being "evidently high"²² (23).

2.2.6. Syntactic Analysis Based on Transition Network Grammars

Woods has developed a grammar described in terms of an augmented transition network. A recursive transition network is a directed graph with labelled states and arcs. The labels on the arcs may be state names or terminal symbols. An arc labelled with a state name produces the following action: the state at the end of the arc is stored in a pushdown stack and control passes to the state indicated by the arc. Control is passed back to the first state by popping the stack. This network is augmented by adding to each arc of the transition network an arbitrary condition which must be satisfied in order for the arc to be followed and a set of structure-building actions to be executed if the arc is followed. The algorithm accepts as input words of a sentence along with the grammatical class of each word of the sentence. The algorithm produces a labelled bracketing of the phrases of the sentence, an identification of the subject, object, verb and an assignment of the sentence type (e.g., interrogative or declarative). The transition

-
22. In order to determine the correct grammatical class of a word, two criteria were used: first, the classification noun is used to signify both nouns and adjectives, and second, the correct classification must correspond to that given by either the Oxford Universal Dictionary or the Merriam-Webster Dictionary.

network model requires about the same (or more) processing time as predictive analyzers require.²³ The accuracy that the augmented transition network attains is not available (24).

2.3. Summary

In the reviews presented above of research concerned with grammatical class assignment, two facts emerge. First, the more economically feasible approaches to grammatical class assignment have treated language as a relational system (at least implicitly). Second, the more successful approaches have avoided analytical procedures which depend upon large lexicons for much of their information.

The purpose of the work described in this chapter was to test the hypothesis that grammatical class assignment could be effected solely on the basis of a knowledge of the SECONDARY RELATIONS and certain PRIMARY RELATIONS contained in a sentence and of the absolute position of each word in the sentence and of the position of each word relative to the RELATIONS surrounding the word.

3. A Basis for Automated Syntactic Analysis

In the identification of NAMES and RELATIONS, one must begin at the word level of analysis. If we suppose that an analytical procedure accepts as input a continuous string of English text, then as a first step, the individual words of that text must be identified. This identification requires a definition of WORD which I present in Section

23. Woods is apparently referring to the Multiple-Path Syntactic Analyzer of Oettinger and Kuno.

4.2. Once the individual words have been identified, the next step is to categorize them in some prescribed way. Traditionally, the categories have been the grammatical classes or parts of speech. Such categories have been used in this research, but it is important to observe that their definitions are different from the traditional ones. Before presenting these definitions, a brief review of earlier work in this regard is presented (Section 3.1.).

Upon categorization of the words in an English text, it is then possible to use a knowledge of these categories and of their sequential ordering in the text to produce larger aggregates (i.e., either COMPOSITE or COMPLEX NAMES).

The categorization of words is treated in this chapter. The aggregation of these categories into more complex units is dealt with in Chapter IV.

3.1. Defining Grammatical Classes

The words which comprise the vocabulary of a language are traditionally classified into eight parts of speech or grammatical classes. Dionysios Thrax is credited with first proposing eight parts of speech: noun, verb, participle, article, pronoun, preposition, adverb and conjunction (25). The interjection is often added to this list. For English, J. Priestly classified words into the above eight parts of speech in 1761 (26). Some parts of speech are given definitions based upon their lexical meaning. For instance, a noun is defined as the name of a person, place or thing. On the other hand, an adjective is defined as a word which modifies a noun. Such a definition is based upon function.

Gleason (27) has suggested an alternative method of defining grammatical classes. Three criteria are used as a definitional basis. The first criterion consists of a paradigm for each of the four classes: noun, verb, adjective and pronoun. Each paradigm consists of possible phonemic inflectional endings which signal membership in the appropriate class. The second criterion lists words whose membership in one of the four classes is signaled by a change in word form rather than by the addition of inflectional endings (e.g., mouse and mice).²⁴ The third criterion is the syntax of the sentence in which a word occurs. This criterion, Gleason feels, is a less sure basis for word class identification than the first two. Gleason's approach to grammatical class assignment suffers because the definition of the class is not sufficiently precise to serve as a basis for the development of algorithms and because it depends upon a subjective determination of a sentence.

Fries, in contrast with Gleason, has defined the parts of speech in terms of a basic structural frame or pattern (28). Each position within a structural frame is occupied by a particular word class. Any word which can fit into a given position in the frame belongs to the corresponding word class. For illustration, consider the sample frame sentences:

1. The concert was good (always).
2. The clerk remembered the tax (suddenly).

24. These first two criteria have been developed into a computer-based stemming system (29).

3. The team went there.

A word which can replace either "concert," "clerk," "tax" or "team" belongs to Class 1. Words which can replace "was," "remembered" or "went" are members of Class 2. Any word which can replace "good" is placed in Class 3, and words which can replace "always," "suddenly" or "there" are placed in Class 4. These four classes roughly correspond to the traditional classes noun, verb, adjective and adverb. Fries chose not to use these terms because of the confusion that might result from definitional differences.

In addition to the four classes defined above, Fries defined a fifth class of words (Class 5) whose elements serve as structural markers within a sentence. These are frequently referred to as function words. They correspond in part to SECONDARY RELATIONS (Chapter II). The members of this class are most readily defined ostensively since there appear to be so few of them (Fries identified just 154). The class is divided into 15 groups which are distinguished by using the same structural criteria as are used for the first four classes. Some of the larger groups within this class are the auxiliary verbs, conjunctions, prepositions, relative pronouns and determiners.

There are significant differences between the first four classes and Class 5. For the first four classes, ~~lexical~~ meanings are easily separable from structural meanings. For function words (Class 5) no clear distinction can be made, perhaps because such words may have no lexical meanings. For instance, while one can detect differences in meaning among the prepositions at, by, for and from, clear distinctions cannot be made

without reference to context.

Another important difference between the first four classes and Class 5 is that whereas the first four classes are open ended, Class 5 appears to be a closed class. As mentioned above, Fries identifies only 154 members of class 5, while there are many thousands of words in each of the other four classes. Furthermore, in an analysis of 1000 words of text, Fries found that function words accounted for almost a third of the total number of word occurrences. And in a more recent analysis of more than 1,000,000 words of text, Kucera and Francis have found that these words account for nearly 46% of the total number of words in the text studied (30).

Finally, it is easy to show that in order to understand certain structural signals within English text, Class 5 words must be known as items. For instance, the two sentences

The boys and the leaders were invited.

The boys of the leaders were invited.

may be analyzed to yield

Class-5 Class-1 Class-5 Class-5 Class-1 Class-5 Class-2

Class-5 Class-1 Class-5 Class-5 Class-1 Class-5 Class-2

The sentences are therefore indistinguishable on this basis. In fact, the only way in which a structural distinction may be made between the two sentences is to know the words and and of as items. In other words, the relationship between boys and leaders is established by the specific items and and of.

Another good example of the structural information provided by Class-5 words is found in the poem of the Jabberwocky:

Twas brillig, and the slithy toves

Did gyre and gimble in the wabe;

All mimsy were the borogoves,

And the mome raths outgrabe

Alice says that "Somehow, it seems to fill my head with ideas - only I don't exactly know what they are!"(31) The ideas which Alice derives from the poem are generated by the structural patterns and the under-scored function words (32).

Function words also serve to eliminate ambiguity. Consider the sentence:

Ship sails today (33).

The ambiguity in this sentence can be avoided by adding function words or other structural markers. The above sentence could have any of the following meanings:

The ship sails today.

Ship the sails today.

Shipped sail today.

Ship sailed today.

The use of function words to resolve ambiguity has also been demonstrated by the work Klein and Simmons in sentence generation (34). After incorporating function words in their program, lexical ambiguity was reduced by 90%. Beckman has also demonstrated that the use of function words in English serve the purpose of an error detecting code (35)

3.2. Formal Definition of Function Words

The work of Fries and others has demonstrated that a knowledge of function words and of the ways in which they operate within a sentence provides a concrete basis for the determination of grammatical classes. But to accomplish the development of an algorithm for grammatical class assignment, a rigorous definition of function words is required.

The definition provided by Fries (36), although structural in character, implies through use of the phrase "can replace" that human judgement is involved in determining class membership. Such a definition is unsuited for algorithm development. On the other hand, the definition can be made suitable by only slight modification: if the phrase "can replace" is replaced by "replaces" then we have a rigorous definition not only of the function words but of all the grammatical classes. There is, however, another disadvantage of Fries' definition that is not so easily overcome. By use of sample frames, grammatical class assignment becomes a process of pattern comparison. It has been shown (37, 38) that simple comparison techniques are unfruitful. The development of viable techniques therefore seems to depend upon a prior knowledge of sentence structure. Thus the use of sample frames leads to a circularity in processing at least when the processing is done algorithmically. As a consequence, I have endeavored to obtain definitions of the grammatical classes which do not depend upon reference to any data save that which is available directly from a sentence.

3.2.1. Definition of Function Words

The class of FUNCTION WORDS is a set of words consisting of both

NAMES and RELATIONS. FUNCTION WORDS are defined as comprising all SECONDARY RELATIONS; the AUXILIARYs, MODALs and ADJUNCTs; certain MAIN RELATIONS; and certain NAMES. Each of the constituent subsets is defined ostensively, as follows.

SECONDARY RELATIONS were defined in Chapter II as consisting of 7 subclasses. Certain of these subclasses are now further subdivided. The complete hierarchy of SECONDARY RELATIONS is shown in Figure 3.1. The elements of each terminal class are listed in Table 3.1.

Among the PRIMARY RELATIONS, the class of ALLIED PRIMARY RELATIONS, consisting of the AUXILIARYs, MODALs and ADJUNCTs, is included in the FUNCTION WORD class. In addition, certain MAIN PRIMARY RELATIONS are treated as FUNCTION WORD elements. Table 3.2 contains the elements of PRIMARY RELATION which are members of FUNCTION WORD.

In order to specify what elements of the class NAME are members of FUNCTION WORD, it is necessary to subdivide NAME as shown in Figure 3.2. The elements of the terminal classes, which are listed in Table 3.3, are important structural markers in the identification of the principal classes NAME and RELATION.

Table 3.4 lists all the terminal classes which are included in FUNCTION WORD. The class of FUNCTION WORDs is thus completely defined and its relation to NAMES and RELATIONS is demonstrated. Of course the subclasses of which FUNCTION WORD is comprised are not mutually exclusive, so that it is necessary to distinguish between them on structural grounds. A set of rules necessary for the complete and unambiguous determination of all word classifications is presented in the next section.

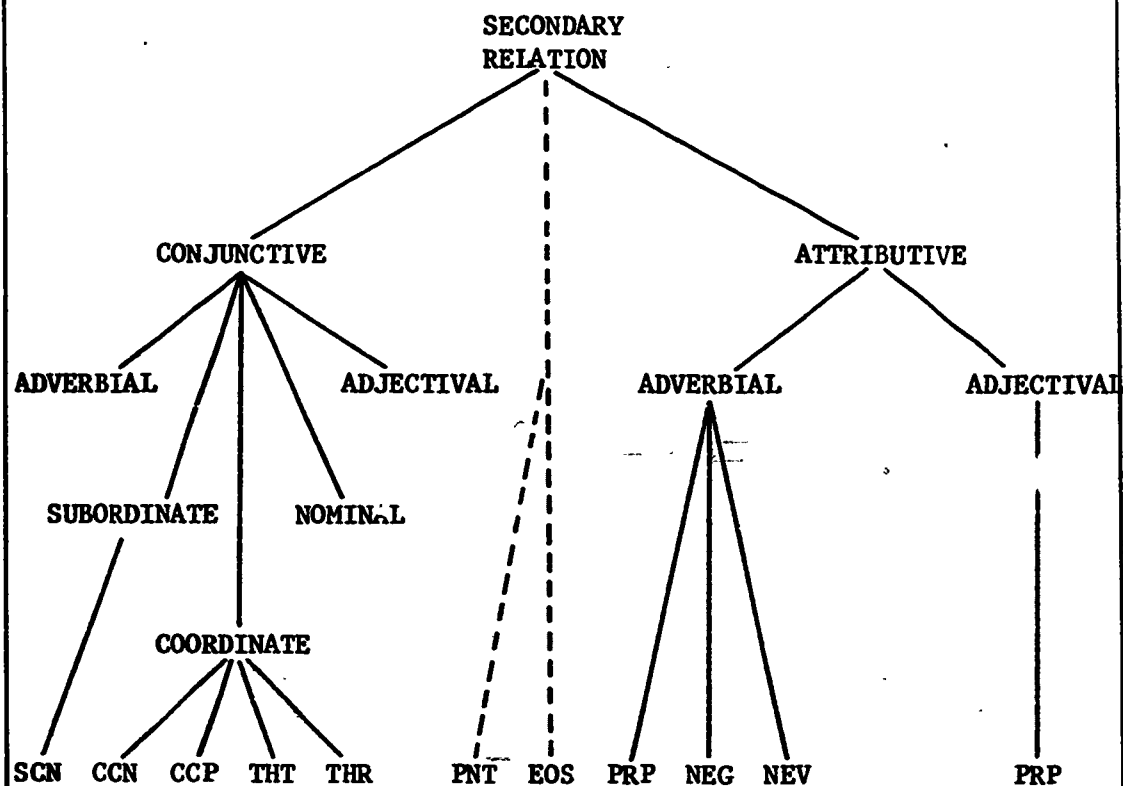


Figure 3.1 Complete hierarchical division of **SECONDARY RELATIONS**. Definition of the terminal classes is given in Table 3.1.

Table 3.1 Specification of the Elements of each Terminal Class in
SECONDARY RELATION.

Class	Elements of the Class
CCN	and but nor not or
CCP	(same as CCN. Distinction made on structural basis. See Rule 71, page 90).
EOS	. ? !
NEG	not
NEV	never
PNT	, ; : ' " ()
PRP	about above according across after along among around at before behind below between beyond by despite down during except for from in inside instead into of off on out outside over through throughout to toward under until up upon with within without
SCN	although however if since than then therefore though thus unless whether yet
THR	there
THT	that
ADJECTIVAL	These classes of CONJUNCTIVE SECONDARY RELATION are structurally defined (see Chapter III, Section 3.4.2.2)
ADVERBIAL	
NOMINAL	

Table 3.2 Specification of the Elements of each Class of PRIMARY RELATION which are Elements of FUNCTION WORD.²⁵

Class	Elements of the Class
AJN	did do does get got let
AUX	am are be been being had has have having is was were
MOD	can cannot could may might must shall should will would
VRB	added appear appeared applied ask asked based became become becomes began believe born bring brought called carried closed come comes concerned consider considered continue continued covered decided described designed determine determined developed done dropped established expect find followed found gave give given gives go going gone happened hear heard held include increased indicated interested involved keep kept knew know known learned led limited lived looked made make makes married meant meet met moved needed obtained opened paid passed placed played prepared provide provided put raised ran reach reached read received related remained remember reported required returned rise said sat saw say says see seem seemed seems seen sent serve set sewed showed shown speak spent started stress suggested take taken tell think told took tried turned understand walk want wanted went worked write written wrote

25. Contractions which contain function words (e.g., I'm) are PRIMARY RELATIONS, but such contractions are not listed here for the sake of simplicity.

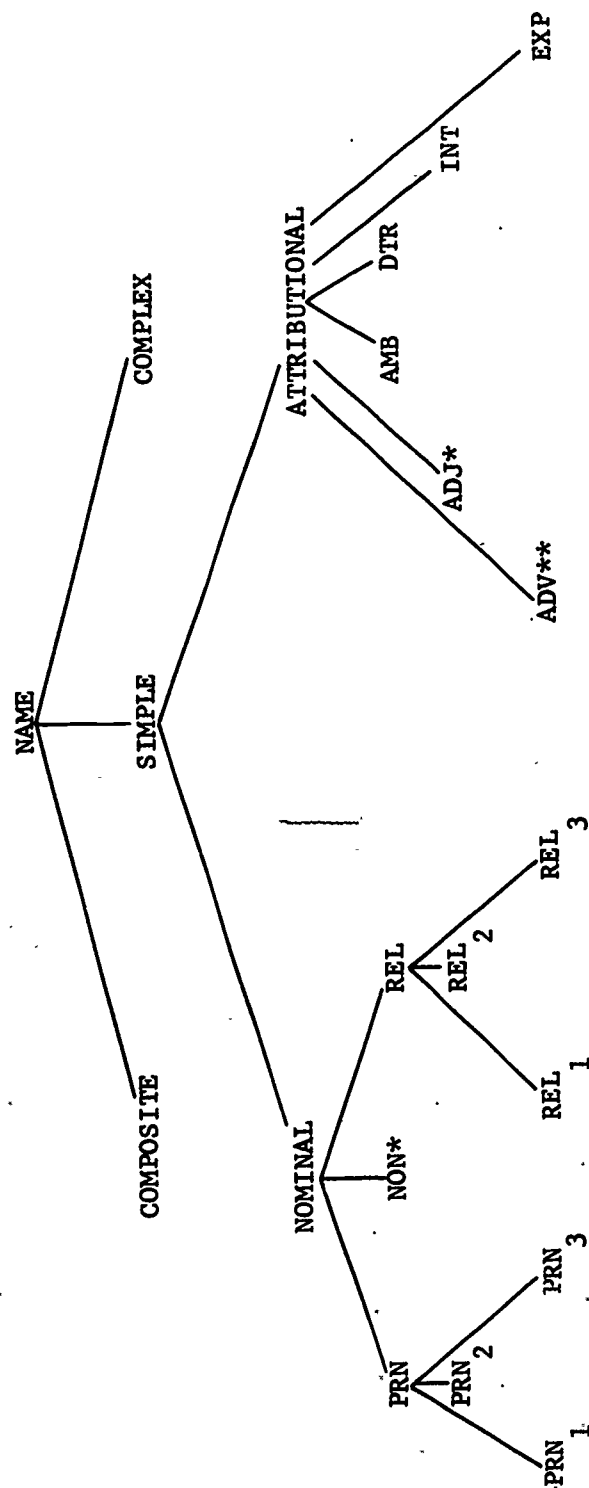


Figure 3.2 Hierarchy of NAMEs showing particularly the terminal classes which are members of FUNCTION WORD. COMPLEX and COMPOSITE NAMES are not members of FUNCTION WORD, nor are those classes marked with an asterisk (*). The double asterisk (**) signifies that those elements of ADV listed in TABLE 3.3 are members of FUNCTION WORD.

Table 3.3 Specification of the Elements of NAME which are Elements of FUNCTION WORD.

Terminal Class	Elements of the Class
ADV	actually again ago ahead almost alone already always away apparently certainly clearly completely daily directly early easily especially even exactly farther finally forward further generally hardly here immediately just later less likely merely near nearly obviously often once only particularly perhaps probably ready really recently simply slowly sometime somewhat soon still suddenly today together too usually
AMB	all another both each eight either enough few five first four her hundred many million more most much neither one ones other own same second several six some ten these this those thru two whole
DTR	a an every his its my our the third your
EXP	oh well
INT	rather quite very
PRN ₁	anyone he him I it me none others she them they thing things us you
PRN ₂	anything everything nothing something
PRN ₃	herself himself itself myself themselves yourself
REL ₁	what whatever who whom
REL ₂	which whose
REL ₃	how when where while why

Table 3.4 Composition of the Class of FUNCTION WORDS.

MAJOR CLASS	TERMINAL CLASSES
NAME	ADJ ADV AMB LTR---EXP INT NON PRN ₁ PRN ₂ PRN ₃ REL ₁ REL ₂ REL ₃
SECONDARY RELATION	CCN CCP EOS NEG NEV PNT PRP SCN THR THR
PRIMARY RELATION	ADJ AUX MOD VRB

3.2.2. Rules for Determination of Grammatical Classes

The FUNCTION WORDS together with the set of rules given below constitute necessary and sufficient conditions for the unambiguous determination of all grammatical classes present in an English sentence. The principal classes to be determined are the three subclasses of NAME: NON, ADJ and ADV; and the three subclasses of PRIMARY RELATION: VRB, INF (INFINITIVAL) and PTC (PARTICIPLE/L). The rules to be described constitute a set of context sensitive productions which are most easily written by means of a simple notation. In addition to the names of the terminal classes given in Table 3.4, the symbols listed in Table 3.5 will be used. A simple example will serve to illustrate the use of the notation. Consider the rule:

... THR XXX ... = ... THR VRB ...

and its interpretation: in a sentence, if the FUNCTION WORD subclass THR is immediately followed by an unclassified word, then that word is assigned to the class VRB.

3.2.2.1. Rules for the Class DTR

Words in this class invariably initiate NOMINAL PHRASES. However, the termination of a NOMINAL PHRASE is not so easily recognized. This problem is illustrated by the following example. Given the "sentence"

DTR XXX XXX PRP DTR DTR XXX EOS

one might be tempted to write a rule which would yield

DTR ADJ NON PRP DTR DTR NON EOS

The definition of SENTENCE given in Chapter II (p. 36) is, however, not satisfied, so that some sort of revision of the initial assignments is

Table 3. . Definition of Notation Symbols used in the Rules for Grammatical Class Assignment.

Symbol	Significance
XXX	any element of a sentence that has not been classified
ZZZ	any element of a sentence that has already been classified (<u>i.e.</u> , a generic class)
~	logical <u>not</u>
	logical <u>or</u>
=	yields
...	elements of unspecified type may be present
'WORD'	any inflected form of the word enclosed in the quote marks
'word'	precisely the word enclosed in the quote marks
XXX'ing'	an unclassified word ending in 'ing'
XXX ⁿ	element repeated n times
()	used to enclose a series of alternatives

necessary. In this instance it is possible to alter the assignments so that

DTR NON VRB PRP DTR DTR NON EOS

is obtained. I have taken precisely this approach in the present research.

Rules involving the class DTR are designed to identify the start of a NOMINAL PHRASE, but they ignore the termination problem. Inaccuracies which this approach brings about are corrected in later stages of the analysis. The rules involving the class DTR are:

Rule 1:

... DTR XXX ... = ... DTR NON ...

Rule 2:

... DTR XXXⁿ ... = ... DTR ADJⁿ⁻¹ NON ... (n ≥ 1)

Rule 3:

... DTR (INT|ADV) XXXⁿ ... = ... DTR (INT|ADV) ADJⁿ⁻¹ NON (n ≥ 1)

Rule 4:

... DTR XXX XXX'ing'ZZZ ... =
... DTR NON PTC ZZZ ...

Rule 5:

... DTR XXXⁿ XXX'ing'ZZZ ... = ... DTR ADJⁿ⁻¹ NON PTC ZZZ ... (n ≥ 1)

Rule 6:

... DTR VRB ... =
... DTR ADJ ...

Rule 7:

... DTR ZZZ ... \Rightarrow
 ... PRN ZZZ ...

3.2.2.2. Rules for the Class AMB

The class AMB consists of words which may belong either to the class DTR or to the class PRN depending upon context. Three simple rules serve to distinguish between the two classes.

Rule 8:

... AMB ZZZ ... \Rightarrow ... PRN ZZZ ...

Rule 9:

... AMB XXX ... \Rightarrow ... DTR XXX ...

Rule 10:

... (AMB|PRN) 'own' ... \Rightarrow ... PRN VRB ...

Note that the last rule also distinguishes 'own' as a member of VRB.

3.2.2.3. Rules for the Class PRN

The two subclasses of PRN, PRN₁ and PRN₂, are helpful in identifying the class VRB. Words in the class PRN₃ frequently are followed by the class ADJ. The following rules are based upon these classes.

Rule 11:

... PRN₁ XXX ... \Rightarrow
 ... PRN₁ VRB ...

Rule 12:

$$\dots \text{XXX} (\text{PRN}_1 | \text{PRN}_2) \dots \Rightarrow \dots \text{VRB} (\text{PRN}_1 | \text{PRN}_2)$$
Rule 13:

$$\dots \text{PRN}_2 \text{XXX} (\text{AUX} | \text{VRB}) \dots \Rightarrow \dots \text{PRN}_2 \text{ADJ} (\text{AUX} | \text{VRB}) \dots$$
Rule 14:

$$\dots \text{PRN}_2 \text{XXX ZZZ} \dots \Rightarrow \dots \text{PRN}_2 \text{VRB ZZZ} \dots$$
Rule 15:

$$\dots \text{PRN}_2 \text{XXX XXX} \dots \Rightarrow \dots \text{PRN}_2 \text{ADJ VRB} \dots$$
Rule 16:

$$\dots \text{PRN}_3 \text{XXX XXX} \dots \Rightarrow \dots \text{PRN}_3 \text{ADJ VRB} \dots$$
Rule 17:

$$\dots \text{PRN}_3 \text{XXX} \dots \Rightarrow \dots \text{PRN}_3 \text{VRB} \dots$$
3.2.2.4. Rules for the Class INT

The class INT gives rise to a single rule.

Rule 18:

$$\dots \text{INT XXX} \dots \Rightarrow \dots \text{INT ADJ} \dots$$
3.2.2.5. Rules for the Class REL

The class REL is divided into three subclasses, REL_1 , REL_2 and REL_3 .

The rules involving these classes are as follows.

Rule 19:

$$\dots \text{'which'} (AMB|PRN|PRP) \dots \Rightarrow \dots REL_2 (AMB|PRN|PRP) \dots$$
Rule 20:

$$\dots REL_1 XXX \dots \Rightarrow \dots REL_1 VRB \dots$$
Rule 21:

$$\dots REL_2 XXX^n \dots \Rightarrow \dots REL_2 ADJ^{n-1} NON \dots \quad n \geq 1$$

If EOS = ? then,

Rule 22:

$$\dots REL_3 XXX \dots \Rightarrow \dots REL_3 VRB \dots$$

otherwise,

Rule 23:

$$\dots REL_3 XXX'ing' \dots \Rightarrow \dots REL_3 PTC \dots$$
Rule 24:

$$\dots REL_3 XXX \dots \Rightarrow \dots REL_3 NON \dots$$
Rule 25:

$$\dots REL_3 XXX XXX \dots \Rightarrow \dots REL_3 NON VRB \dots$$

Rule 26:

$$\dots \text{REL}_3 \text{XXX}^n \dots \Rightarrow \dots \text{REL}_3 \text{ADJ}^{n-2} \text{NON VRB} \dots \quad n \geq 2$$

In general,

Rule 27:

$$\dots \text{REL EOS} \Rightarrow \dots \text{ADV EOS}$$

3.2.2.6. The Class ADV

Words in this class cannot be reliably identified by any rules so far considered. For this reason, a list of words in the class ADV has been incorporated in the class FUNCTION WORD so that the rules which have been developed may prove to function more reliably. The words in this group are given in Table 3.3.

3.2.2.7. Rules for the Classes AJN, AUX and MOD

The elements of these classes are relatively few in number and the classes are reliable predictors of juxtaposed classes. Therefore a relatively large number of rules has been devised for these classes. The form of the rules differs in some instances depending upon the value of EOS. Rules involving the class AUX follow.

Rule 28:

$$\dots \text{'BE'} (\text{ADV}|\text{NEG}|\text{NEV}) \text{'being'} \Rightarrow \dots \text{'BE'} (\text{ADV}|\text{NEG}|\text{NEV}) \text{PTC} \dots$$

Rule 29:

$$\dots \text{'BE'} \text{'being'} \dots \Rightarrow \dots \text{'BE'} \text{PTC} \dots$$

Rule 30:

... 'being' XXX ... \Rightarrow
 ... (PTC|AUX) ADJ ...

Rule 31:

... 'being' XXXⁿ ... \Rightarrow
 ... (PTC|AUX) ADJⁿ⁻¹ NON $n > 1$

Rule 32:

... 'BE' (XXX'ing' |XXX'ed')... \Rightarrow
 ... AUX VRB ...

Rule 33:

... 'BE' ADV (XXX'ing' |XXX'ed')... \Rightarrow
 ... AUX ADV VRB ...

Rule 34:

... 'BE' XXX ... \Rightarrow
 ... 'BE' ADJ ...

Rule 35:

... 'BE' (INT|ADV) XXX ... \Rightarrow
 ... AUX (INT|ADV) ADJ

Rule 36:

... 'BE' (INT|ADV) XXXⁿ ... \Rightarrow
 ... 'BE' (INT|ADV) ADJⁿ⁻¹ NON $n \geq 1$

Rule 37:

... 'BE' XXXⁿ ... \Rightarrow
 ... 'BE' ADJⁿ⁻¹ NON ... $n \geq 1$

Rule 38:

... 'BE' 'having' ... =
... AUX VRB ...

Rule 39:

... 'BE' (ADV|NEG|NEV) 'having' ... =>
... AUX (ADV|NEG|NEV) VRB

Rule 40:

... 'having' XXX'ed' ... =>
... PTC PTC

Rule 41:

... 'having' 'been' XXX'ed' ... =
... PTC PTC PTC ...

Rule 42:

... 'having' ... ⇒
... PTC ...

Rule 43:

... 'HAVE' XXX'ed'... =>
... 'HAVE' VRB

Rule 44:

... 'HAVE' XXX ... =
... 'HAVE' NON ...

Rule 45:

$$\dots \text{'HAVE'} \text{XXX}^n \dots \Rightarrow \dots \text{'HAVE'} \text{ADJ}^{n-1} \text{NON}^* \dots \quad n \geq 1$$

If EOS = ?, then

Rule 46:

... AUX XXX (AUX|VRB) ... \Rightarrow
 ... AUX NON (AUX|VRB)

Rule 47:

... AUX XXXⁿ (AUX|VRB) ... \Rightarrow
 ... AUX ADJⁿ⁻¹ NON (AUX|VRB) ...

Rule 48:

... AUX XXX ... \Rightarrow
 ... AUX NON

Rule 49:

... AUX XXX XXX ... \Rightarrow
 ... AUX NON VRB ...

Rule 50:

... AUX XXXⁿ ... \Rightarrow ... AUX ADJⁿ⁻¹ NON VRB $n \geq 2$

The following rules involve the class MOD.

Rule 51:

... MOD XXX ... \Rightarrow
 ... MOD VRB ...

Rule 52:

... MOD ADV XXX ... \Rightarrow
 ... MOD ADV VRB ...

Rule 53:

... (AUX|VRB) ... 'can' (DTR|PRP|PRN) ... \Rightarrow
 ... (AUX|VRB) ... VRB (DTR|PRP|PRN) ...

Rule 54:

... (PRP|AUX|DTR) ('can'|'may'|'will') ... ⇒
... (PRP|AUX|DTR) NON ...

Rule 55:

... ('can' | 'may' | 'will') (PRP | DTR) ... =
... NON (PRP | DTR) ...

The following rules involve the class AJN.

Rule 56:

... ('get' | 'gets') (XXX'ing' | XXX'ed')... =
... AUX VRB ...

Rule 57:

... ('keep' | 'keeps' | 'kept') XXX'ing' ... =>
... AUX VRB

Rule 58:

... ('let' | 'lets') XXX ... =>
... AUX VRB ...

Rule 59:

... 'let' XXX PRP ... =>
... AUX VRB PRP

Rule 60:

... 'let' XXX XXX ... ⇒
... AUX NON VRB ...

Rule 61:

... 'let' XXXⁿ ... ⇒
... 'let' ADJⁿ⁻² NON VRB n > 2

Rule 62:

... 'let' DTR XXX XXX ... =
 ... 'let' DTR NON VRB ...

Rule 63:

... 'let' DTR XXXⁿ ... =
 ... 'let' DTR ADJⁿ⁻² NON VRB ... $n \geq 2$

Rule 64:

... 'let' PRN XXX ... =
 ... 'let' PRN VRB

Rule 65:²⁶

... ('did' | 'does' | 'do') NEG XXX ... =
 ... ('did' | 'does' | 'do') NEG VRB ...

Rule 66:

... ('did' | 'does') XXX ... =
 ... AUX VRB ...

Rule 67:

... ('did' | 'does') XXX XXX ... =
 ... AUX VRB NON ...

Rule 68:

... ('did' | 'does') XXXⁿ ... =
 ... ('did' | 'does') VRB ADJⁿ⁻² NON ... $n \geq 2$

26. The forms 'did' NEG; 'does' NEG and 'do' NEG are equivalent to 'didn't', 'doesn't' and 'don't', respectively.

Rule 69:

EOS 'DO' XXX ... =
EOS 'DO' VRB ...

3.2.2.8. The Class VRB

The PRIMARY RELATIONS are the most difficult to identify reliably. For this reason a number of the most commonly occurring members of this major class are included in the FUNCTION WORDS as the class VRB. The VRB class gives rise to no rules, but assists in the correct operation of other rules. The elements of VRB are listed in Table 3.2.

3.2.2.9. Rules for the Classes CCN, CCP and SCN

The elements of CCN and CCP are identical. Initially the elements which comprise these classes are identified as elements of CCN. The rules presented in this section identify the instances in which a word is an element of CCP. In later processing, the elements of CCN are used to identify clause boundaries. Some elements of SCN are also elements of VRB and PRP. The rules presented here identify these occurrences and reclassify the SCN element accordingly. The elements of CCN, CCP and SCN are listed in Table 3.1.

Rule 70:

... NON CCN (ME|HE|THEM|SHE|WE) ... =
... NON CCN PRN

Rule 71:

... ADJ CCN XXX (EOS|PNT|CCN|SCN|THT) ... =
... ADJ CCP ADJ ZZZ ...

Rule 72:

... ADJ CCN (ADV|INT) XXX (EOS|PNT|CCN|SCN|THT) ... =
 ... ADJ CCP (ADV|INT) ADJ ...

Rule 73:

... (AUX|VRB) CCN XXX (DTR|PRN|PRP|EOS|PNT|CCN|SCN|THT) ... =
 ... (AUX|VRB) CCP VRB ZZZ ...

Rule 74:

... NON CCN XXX (EOS|PNT|CCN|SCN|THT) ... =
 ... NON CCP NON ZZZ ...

Rule 75:

... SCN XXX'ing' ... =
 ... SCN PTC ...

Rule 76:

... SCN EOS =
 ... ADV EOS

Rule 77:

... ('like'|'since', ... (AUX|VRB) ... (EOS|PNT|REL|CCN|SCN|THT) ... =
 ... PRP ... (AUX|VRB) ... ZZZ

Rule 78:

... ('like'|'to') INF ... =
 ... VRB PRP INF ...

3.2.2.10. Rules for the Class THR

The class THR while sometimes having the role of an ADV, generally initiates a clause and precedes the class VRB. If THR follows VRB or if it precedes a CONJUNCTIVE SECONDARY RELATION, it is classified as ADV.

Rule 79:

... THR XXX ... \Rightarrow
... THR VRB ...

Rule 80:

... (VRB|AUX) THR ... \Rightarrow
... (AUX|VRB) ADV ...

Rule 81:

... THR PRP ... \Rightarrow
... ADV PRP ...

3.2.2.11. Rules for the Class THT

The class contains the single element "that." The property which makes this element unique is that the element may belong to either the AMB class, the SCN class or the CONJUNCTIVE SECONDARY RELATIONS. Rules which apply to THT are similar to the AMB rules; however, THT is used in the later stages of MYRA (see Section 3.2.3.) to identify clause boundaries.

Rule 82:

... THT ZZZ ... \Rightarrow ... PRN ZZZ ...

Rule 83:

... THT XXX ... \Rightarrow ... DTR XXX ...

3.2.2.12. Rules for the Class PRP

Words in this class are used to identify NOMINAL PHRASEs, and elements of INF, and PTC. The elements of PRP are listed in Table 3.1.

Rule 84:

... 'to' XXX ... \Rightarrow
 ... PRP INF ...

Rule 85:

... 'to' ADV XXX ... \Rightarrow
 ... PRP ADV INF ...

Rule 86:

... 'to' XXXⁿ ... \Rightarrow ... PRN ADJⁿ⁻¹ NON $n \geq 1$

Rule 87:

... PRP XXX'ing' ... \Rightarrow
 ... PRP PTC ...

Rule 88:

... PRP XXX XXX (AUX|VRB) ... \Rightarrow
 ... PRP NON NON (AUX|VRB) ...

Rule 89:

... PRP XXXⁿ (AUX|VRB) ... \Rightarrow ... PRP ADJⁿ⁻² NON NON (AUX|VRB) ... $n \geq 2$

Rule 90:

... PRP XXX ... \Rightarrow
 ... PRP NON ...

Rule 91:

... PRP XXXⁿ ... \Rightarrow ... PRP ADJⁿ⁻¹ NON $n \geq 1$

Rule 92:

$$\dots \text{PRP (INT ADV) XXX}^n \dots \Rightarrow \dots \text{PRP (INT ADV) ADJ}^{n-1} \text{NON} \quad n \geq 1$$
Rule 93:

$$\dots \text{PRP (EOS PNT)} \dots \Rightarrow \dots \text{ADV (EOS PNT)} \dots$$
Rule 94:

$$\dots \text{XXX'ing' PRP} \dots \Rightarrow \dots \text{PTC PRP} \dots$$
3.2.2.13. Rules for the Class NEV

This class is important because of its reliability in signalling the presence of a PRIMARY RELATION. NEV contains the single element "never."

Rule 95:

$$\dots \text{NEV XXX} \dots \Rightarrow \dots \text{NEV VRB} \dots$$
Rule 96:

$$\dots \text{NEV ADV XXX} \dots \Rightarrow \dots \text{NEV ADV VRB} \dots$$
Rule 97:

$$\dots \text{NEV XXX'ing'} \dots \Rightarrow \dots \text{NEV PTC} \dots$$
3.2.2.14. Rules for the Class NEG

NEG contains the single element "not." This element can be classed as either a CCN or as a component of a PRIMARY RELATION. The following

rules differentiate these uses.

Rule 98:

... AUX NEG XXX ... =
... AUX NEG VRB ...

Rule 99:

... MOD NEG XXX ... =
... MOD NEG VRB ...

Rule 100:

... NEG XXX 'ing' ... =
... NEG PTC ...

Rule 101:

... NEG ... =
... CCN ...

3.2.3. Sufficiency of the Rules

In Section 3.2.2. it was stated that a knowledge of FUNCTION WORDS, together with the rules just described constitute necessary and sufficient conditions for the unambiguous determination of all grammatical classes as defined in this research. I believe these are necessary conditions: all elements of FUNCTION WORD are words that can reasonably be expected to occur in an English sentence and which serve the relational purpose defined for FUNCTION WORD; and the rules have all been established on the basis of patterns actually observed in English text (39). The question is therefore one of sufficiency.

It is not possible to say with certainty that FUNCTION WORD is complete, therefore it cannot be said to be sufficient on that basis. However, experimentation has shown (see Section 4) that for practical purposes it appears sufficient. Thus the question of sufficiency rests with the rules. Here it is easy to show that they are insufficient since a sentence that contains no element of FUNCTION WORD cannot be associated with any rule so far given. Furthermore, there is ample experimental evidence to show that even if this case were eliminated from consideration, there would yet be found structural patterns which the above rules do not account for. One must conclude therefore that the rules are insufficient. Can they be made sufficient? The answer to this question is a qualified "yes." The qualification is that the rules can be made sufficient in the sense that every element of every sentence can be classified. The question of accuracy of classification will be dealt with later.

Two types of rules need to be added to those already presented in order to complete the sufficiency condition. The first type is concerned with the classification of sentence elements which have not yet been classified. The second type deals with reclassification of previously classified elements so as to satisfy the definition of SENTENCE given earlier (Section 3.3.3.). These rules follow:

Rule 102:

... XXXⁿ ... ⇒ ... VRB ...

n = 1

Rule 103:

$$\dots \text{XXX}^n \dots \Rightarrow \dots \text{NON VRB} \dots \quad n = 2$$
Rule 104:

$$\dots \text{XXX}^n \dots \Rightarrow \dots \text{ADJ}^{n-2} \text{NON VRB} \dots \quad n > 2$$
Rule 105:

$$\begin{aligned} \dots \text{XXX}^n \text{DTR ADJ}^m \text{NON ADJ}^p \text{NON} \dots &\Rightarrow & n = 0 \\ \dots \text{XXX}^n \text{DTR ADJ}^{m-1} \text{NON VRB ADJ}^p \text{NON} \dots && m > 0 \\ && p \geq 0 \end{aligned}$$
Rule 106:

$$\begin{aligned} \dots \text{XXX}^n \text{NON ADJ}^p \text{NON} \dots &\Rightarrow & n = 0 \\ && p > 1 \end{aligned}$$
Rule 107:

$$\begin{aligned} \dots \text{XXX}^n \text{ADJ}^m \text{NON ADJ}^p \text{NON} \dots &\Rightarrow & n = 0 \\ && m > 0 \\ \dots \text{XXX}^n \text{ADJ}^{m-1} \text{NON VRB ADJ}^p \text{NON} \dots && p \geq 0 \end{aligned}$$

If none of the above rules apply the following rules are applied in a left-to-right manner until a VRB assignment is made or until the end of the sentence is encountered.

Rule 108:

$$\dots \text{DTR ADJ}^n \text{NON} \dots \Rightarrow \dots \text{DTR ADJ}^{n-1} \text{NON VRB} \dots \quad n > 0$$
Rule 109:

$$\dots \text{ADJ}^n \text{NON} \dots \Rightarrow \dots \text{ADJ}^{n-1} \text{NON VRB} \dots \quad n \geq 0$$
Rule 110:

$$\dots \text{NON} \dots \Rightarrow \dots \text{VRB} \dots$$

4. MYRA - A Program for Grammatical Class Assignment

A program, called MYRA, has been developed which accepts English text as input and produces as output each word of text together with the name of the class to which it belongs. The class of FUNCTION WORDs and the rules described in Section 3 form the basis for the program.

MYRA is described in this section. Experimental results involving MYRA are given in Section 5.

4.1. General Description of MYRA

MYRA has been written in PL/I, and compiled using the F-level compiler of I.B.M. operating on an IBM System 370-165 (PHENIX XV, HASP II). MYRA is capable of processing text at a rate of 13,500 words of text per minute of C.P.U. usage. Approximately 126,000 bytes of storage are required for the program and working storage.

The heart of MYRA is a dictionary and a set of rules. Each of these is discussed in the following paragraphs.

4.1.1. The Dictionary

The class of FUNCTION WORDs defined in Section 3.2.1. is incorporated in a dictionary. The dictionary is ordered first by the length (number of characters) of its elements. If several elements have the same length, then each of these is ordered alphabetically. Such an ordering facilitates the dictionary look-up process (see Section 4.2.1., below). Associated with each element of the dictionary is a numerical code which identifies the class to which the element belongs and serves as a unique NAME for the element. These data are

helpful in updating the dictionary. A program has been written to handle dictionary updates. It produces a dictionary acceptable to MYRA.

Two versions of the dictionary have been established for test purposes (see Section 5). One, the limited version, consists of all the terminal classes listed in Table 3.4, except ADV and VRB, and contains 217 elements. The second, called the extended dictionary, consists of all the terminal classes of Table 3.4. and contains 431 elements. The extended dictionary is included as Appendix A. The distribution of elements by length is shown in Figure 3.3.

4.1.2. The Rules

MYRA consists basically of a series of PL/I statements which correspond to the rules given in Section 3.2.2. The "execution" of these statements results in the classification of the words of a sentence.

4.2. Operation of MYRA

MYRA accepts English text as input in a continuous string without any prior formatting or marking. This fact has two implications. First, MYRA doesn't "know" that the input is English text, but the text will be processed as though it were. Second, MYRA must break the string up into individual words. A WORD is defined as any string of characters bounded by blanks, except that the elements of the classes PNT and EOS are isolated as individual words and labelled as members of the appropriate class. Hence, a string such as

becomes,

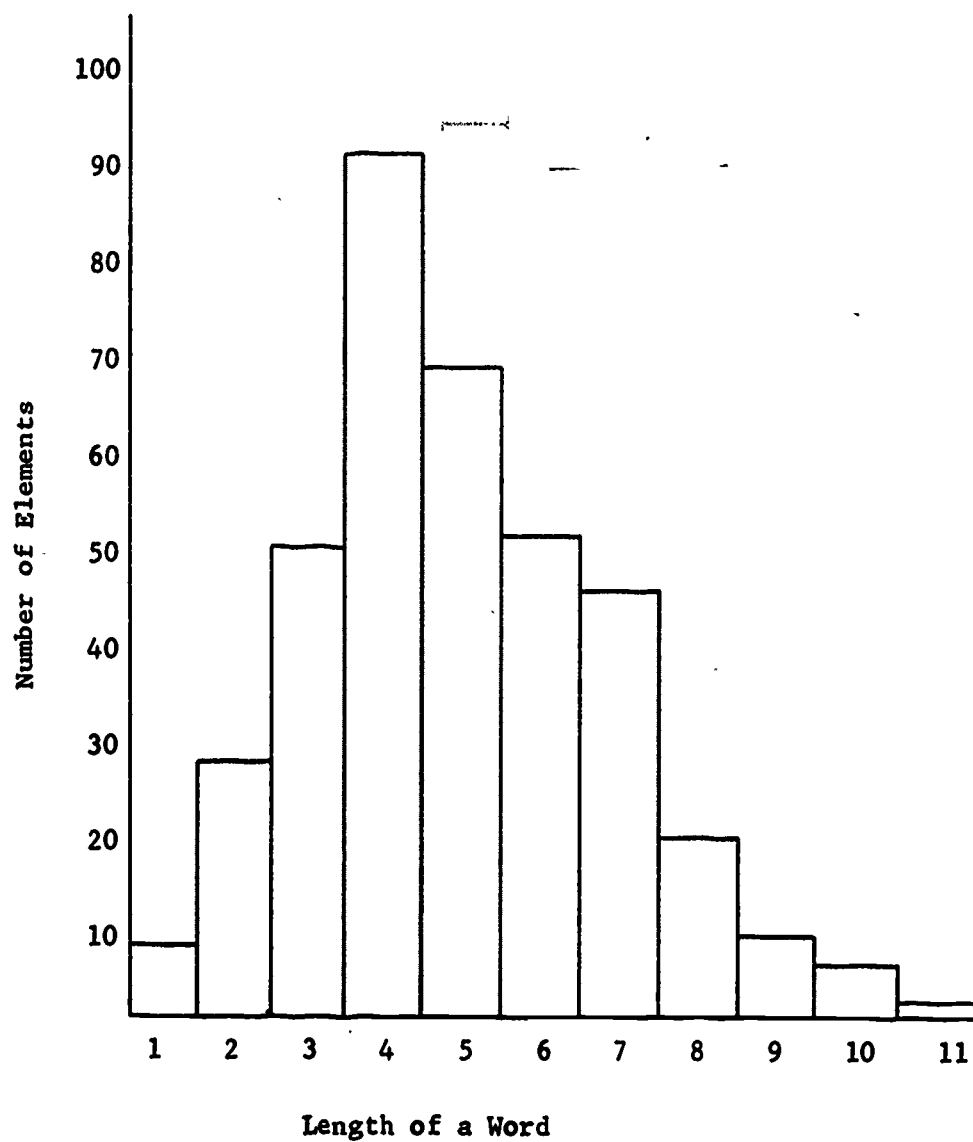


Figure 3.3 A bar graph of the number of elements of a given length in the extended dictionary.

is separated into two WORDs:

becomes

and

Once WORDs have been isolated, a preliminary SENTENCE boundary must be established. Any element of EOS serves this purpose. MYRA is then ready to effect grammatical class assignments.

MYRA operates on the input text, partitioned as outlined above, in three stages.

4.2.1. Dictionary Look-Up

In the first stage, the individual words, exclusive of those in PNT or EOS are looked up in the dictionary. The input word is matched only against those dictionary elements of the same length. If a match occurs, the code for the dictionary element is entered into a vector which corresponds with the sequence of words between elements of EOS. For instance, the input string

The mouse ate the cheese.

would have already been partitioned as

The/mouse/ate/the/cheese/.

and a corresponding vector

. XXX XXX XXX XXX XXX EOS

set up.²⁷ If the first word of the string is looked up in the dictionary, a match will occur and the corresponding code retrieved. When all WORDs have been looked up (terminating with EOS), the program begins stage two. For our example, the vector resulting at the end of stage one will be

DTR XXX XXX DTR XXX EOS

4.2.2. Application of Rules 1-101

In the second stage, MYRA applies Rules 1-101. Application is signalled by the presence in the vector of an element of FUNCTION WORD or of some other element(s) already classified, so that only rules which can reasonably be expected to produce classifications are applied. In our example, we have at first a rule for DTR which can be applied (Rule 2). Its application yields

DTR ADJ NON DTR XXX EOS

Moving to the right in the vector, we see that a second rule for DTR is called for (Rule 1). Its application yields

DTR ADJ NON DTR NON EOS

Note that in stage two MYRA operates on the vector corresponding to the input string and does not process the input except in the instance of those rules that require a particular inflectional ending (i.e., 'ing').

27. I shall use the codes previously defined for illustration, rather than the numerical codes employed in MYRA.

At the end of stage two, the vector may or may not be complete. It is completed and verified in stage three.

4.2.3. Application of Rules 102-110

In stage three MYRA first classifies any previously unclassified WORDs. This is done by application of Rules 102-104. Finally, the vector is checked to see if a PRIMARY RELATION is present. (If the class AUX is not part of a COMPOSITE PRIMARY RELATION, AUX is reclassified as MAIN). If no PRIMARY RELATION is present, then Rules 105-110 are applied to reclassify elements of the vector so that a PRIMARY RELATION is included. The analysis is thus completed and the results are output. In our example, at the end of stage two the vector contained no PRIMARY RELATION. Application of Rule 108 causes the sequence

DTR ADJ NON ...

to be transformed to

DTR NON VRB ...

so that the final vector would be

DTR NON VRB DTR NON EOS

The application of MYRA to several English texts and the results obtained are described in the next section.

5. Experimental Testing of MYRA

In order to determine the efficacy of MYRA, three bodies of text, derived from

- a) The Need for More Precise Definition of "Algorithm" by
B. A. Trakhtenbrot
- b) The Old Man and the Sea by Ernest Hemingway
- c) The Clavichord and How to Play It by M. Halford

and totaling about 6000 words were used. These texts were processed using both the limited and extended dictionaries (Section 4.1.1.).

The results were then analyzed manually in order to determine the accuracy of the classifications. Wherever an error was found, its cause was identified (as far as possible). The accuracy of the classification was based upon my own intuitive knowledge of English. Hence, I set myself up as the standard against which the results of MYRA were evaluated.

The results of this analysis are presented in Tables 3.6 and 3.7. Using the limited dictionary, MYRA produced results that were 91% accurate on average using the evaluation criterion mentioned above. With the extended dictionary, MYRA achieved an average accuracy of 94%. More detail concerning the evaluation of MYRA will be found in Appendix D. Sample output from MYRA is shown in Figure 3.4. Complete output for the second body of text (The Old Man and the Sea) is included as Appendix C.

Table 3.6 Error Analysis of the Results Produced by MYRA Using the Limited Dictionary.

Text	Overall Accuracy	% of Errors by Type			
		NON	VRB	ADJ	OTHER
"...Precise Definition of 'Algorithm'"	92%	28%	35%	10%	11%
<u>The Old Man</u> and <u>the Sea</u>	90%	26%	33%	14%	15%
"The Clavichord..."	92%	25%	28%	21%	10%

Table 3.7 Error Analysis of the Results Produced by MYRA Using the Extended Dictionary.

Text	Overall Accuracy	% of Errors by Type			
		NON	VRB	ADJ	OTHER
"...Precise Definition of 'Algorithm'"	94%	33%	20%	17%	15%
<u>The Old Man</u> and <u>the Sea</u>	95%	33%	27%	24%	7%
"The Clavichord..."	94%	21%	22%	29%	13%

The old man had taught the boy to fish and the boy loved him.

DTR ADJ NON AUX VRB DTR NON PRP VRB CNJ DTR NON VRB PRN EOS

Figure 3.4 Sample output produced by MYRA from the text The Old Man and the Sea.

6. Summary

In Chapter II I have presented a theoretical framework suitable for the syntactic analysis of English text. Based upon this framework, a program, MYRA, has been developed implemented and tested which assigns words in a sentence to their appropriate grammatical classes (i.e., identifies them as NAMES or RELATIONS). The test results have been analyzed and an accuracy of identification of between 91% and 94% has been found. Thus, MYRA has been shown to have a theoretical base, to produce accurate results and to operate at a high rate of speed (at 13,500 words per minute).

The output of MYRA forms the input to procedures described in Chapter IV. Further conclusions which may be drawn from the results so far obtained in this research will be deferred until the last chapter of this dissertation.

References

1. A. Oettinger and S. Kuno, "Multiple-path Syntactic Analyzer," Proceedings of the International Federation of Information Processing Congress, North Holland Publishing Co., Munich, 1962, 306-312.
2. ibid., p. 306.
3. ibid., p. 310.
4. A. M. Zwicky, J. Friedman, B. C. Hall and D. E. Walker, "The MITRE Syntactic Analysis Procedure for Transformational Grammars," Proceedings of the Fall Joint Computer Conference 27 (01), 317-326 (1965).
5. ibid., p. 325.
6. ibid., p. 325.
7. Z. Harris, Mathematical Structure of Language, Interscience Publishers, 1968, pp. 230.
8. N. Sager and M. Salkoff, "Syntactic Analysis of Natural Language," Advances in Computers 08 (00), 1969, 153-188.
9. T. Winograd, "Understanding Natural Language," Cognitive Psychology 3, 1-191 (1972).
10. M. A. K. Halliday, "Notes on Transitivity and Theme in English, Part 1," Journal of Linguistics 3 (1), 37-81 (1967).
11. M. A. K. Halliday, "Functional Diversity in Language as Seen from a Consideration of Modality and Mood in English," Foundations of Language 6, 322-361 (1970).
12. M. A. K. Halliday, "Notes on Transitivity and Theme in English, Part 2," Journal of Linguistics 3 (2), 200 (1967).
13. Winograd, op. cit.
14. D. C. Clarke and R. E. Wall, "An Economical Program for Limited Parsing of English," Proceedings of Fall Joint Computer Conference 307-316 (1965).
15. ibid., p. 314.

16. S. Klein and R. Simmon, "A Computational Approach to Grammatical Coding of English Words," Journal of the ACM 10 (03), 342 (1963).
17. ibid., p. 344.
18. W. Stolz, P. Tannenbaum and P. Carstensen, "A Stochastic Approach to the Grammatical Coding of English," Communication of the ACM 8 (6), 399-405 (1965).
19. ibid., p. 404.
20. J. P. Thorne, P. Bratley and H. Dewar, "The Syntactic Analysis of English by Machine," in Machine Intelligence 3, D. Michie (ed.), American Elsevier Publishing Co., Inc., New York, 1969, pp. 281-309.
21. ibid., p. 307 and p. 303.
22. H. L. Resinkoff and J. L. Dolby, "Automatic Determination of the Parts of Speech of English Words," Proceedings of IEEE 15 (7), 1029 (1963).
23. ibid., p. 1029.
24. W. A. Woods, "Transition Network Grammars for Natural Language Analysis," Communications of the ACM 13 (10), 591-606 (1970).
25. L. Bloomfield, Language, Holt, Rinehart and Winston, New York, 1933, p. 5.
26. J. Priestly, Rudiments of English Grammar, London, 1761, in C. Fries, The Structure of English, Harcourt, Brace & World, Inc., New York, 1952.
27. H. A. Gleason, An Introduction to Descriptive Linguistics, Holt, Rinehart and Winston, New York, 1961.
28. C. Fries, The Structure of English, Harcourt, Brace & World, Inc., New York, 1952, p. 74.
29. T. L. Lowe, "Direct-Access Memory Retrieval Using Truncated Record Names" Software Age, September, 28-33, 1967.
30. H. Kucera and W. R. Francis, Computational Analysis of Present-Day American English, Brown University, Providence, R. I., 1967.

31. L. Carroll, "Through the Looking Glass," from Alice's Adventures in Wonderland and Through the Looking Glass, Airmont Publishing Co., New York, 1965, p. 141.
32. Fries, op. cit., p. 70.
33. ibid., p. 70.
34. D. G. Bobrow, "Syntactic Analysis of English by Computer--A Survey," AFIPS Conference Proceedings 24, 365-387 (1963).
35. P. Beckmann, The Structure of Language: A New Approach, The Golem Press, Colorado, 1972.
36. Fries, op. cit., p. 75.
37. L. L. Earl, Annual Report: Automatic Indexing and Abstracting, Lockheed Missiles and Space Co., Palo Alto, California, 1967 (AD 659 057).
38. L. L. Earl and H. R. Robinson, Automatic Informative Abstracting and Extracting, Lockheed Missiles and Space Co., Palo Alto, California, 1968 (AD 667 473).
39. S. Marvin, J. E. Rush, and C. E. Young, "Grammatical Class Assignment Based on Function Words," from A. D. Berton (ed.) 7th Annual National Colloquium on Information Retrieval, Medical Documentation Service, Philadelphia, Pennsylvania, 1970, pp. 71-81.

CHAPTER IV. IDENTIFICATION OF COMPOSITE AND COMPLEX NAMES

All the innumerable substances which occur on earth--shoes, ships, sealing-wax, cabbages, kings, carpenters, walruses, oysters, everything we can think of--can be analysed into their constituent atoms, either in this or in other ways. It might be thought that a quite incredible number of different kinds of atoms would emerge from the rich variety of substances we find on earth. Actually the number is quite small. The same atoms turn up again and again, and the great variety of substances we find on earth results, not from any great variety of atoms entering into their composition, but from the great variety of ways in which a few types of atoms can be combined--

Sir James Jeans, The Universe Around Us

There are countless ways of writing English sentences. ... But sentences in English have certain elements in common, and when you start to analyze these sentences, you will find that there are a very few basic sentence patterns that all writers use.

Ann Eljenholm Nichols, English Syntax

1. Introduction

The preceding chapter has dealt with procedures for identifying SIMPLE NAMES and RELATIONS, procedures which were based upon a knowledge of a special class of WORDs, called FUNCTION WORDs, and upon a set of rules involving structural patterns. In this chapter I build upon the previous results. Procedures are described whose purpose is the identification of COMPOSITE and COMPLEX NAMES. These procedures are based upon the definition of COMPOSITE NAME and COMPLEX NAME given in Chapter II, and upon structural signals provided by elements of FUNCTION WORD and by the arrangement of SIMPLE NAMES and SIMPLE RELATIONS within a sentence as identified by MYRA. Before describing these

procedures, brief mention must be made of certain related research. In Chapter III, Section 2.1, several projects were reviewed in which either a form of top-down parser (1, 2, 3, 4) or of bottom-up parser (5) was used. The output of these parsers included grammatical class assignment, phrase-type recognition and clause identification. These parsers are similar to one another in that each employs a large dictionary which contains all possible grammatical classes for each lexical entry. The size of the dictionary and the complexity of the procedures result in long processing times. One possible exception to this statement may be afforded by the work of Woods (6). Although extant publications by Woods indicate processing times comparable to those realized for other top-down parsers, unconfirmed reports indicate that Woods has substantially decreased his processing times (7).

One of the projects reviewed in Chapter III not only involved the identification of word classes, but also the identification of nine phrase types (noun, prepositional, pronoun, infinitive, verb, adverbial, post-modifying adjective, present participle and past participle) (8). Stolz, Tannenbaum and Carstensen reported an accuracy of 91% in the application of their procedures to both technical and nontechnical abstracts (8). Clauses were also identified, but the details of clause identification were not given.

In general, previous work on phrase and clause identification has relied heavily on lexical information and it was therefore of interest to test the hypothesis that the desired results could be attained without recourse to extensive dictionaries and with rather minimal

rules. A description of the results of such tests forms the remainder of this chapter.

2. Procedures for the Identification of CLAUSES

In this section is described a prototype system for the identification of the class of COMPLEX NAMES called CLAUSES as they occur in English text. This system is one component of the language analysis system described in this dissertation. An improved version of this prototype is outlined in Section 4.

The set of procedures which identify CLAUSES is called CAP/I (CLAUSE Analysis Procedures/I). Input to CAP/I is the output of MYRA (Chapter III). The delimitation of CLAUSES by CAP/I is based primarily upon structural signals. These signals include the classes of CONJUNCTIVE SECONDARY RELATION, namely SCN, CCN, CCP, THT and THR, as well as the classes PNT, EOS, INF and PTC (see Table 3.1 for definition).

CAP/I operates in two phases. CLAUSES are identified in either phase according to the structure of the SENTENCE which contains them and according to the type of CONJUNCTIVE SECONDARY RELATION (if present) that introduces the CLAUSE. The rules employed in each phase of CAP/I are described below.

2.1. Phase I of CAP/I

In phase I, CAP/I examines successive WORDs of a SENTENCE until one of the elements of CCN, THT, PNT or REL is encountered (see Tables 3.1 and 3.3 for definition of these classes). When one of these RELATIONS is encountered, rules for the particular RELATION are applied to the SENTENCE. The purpose of phase I is to make initial

CLAUSE boundary assignments. Other CLAUSE boundaries are added in phase II.

2.1.1. Rules for the Class CCN

When an element of CCN is encountered, several checks are made to ascertain that the CCN marks a CLAUSE boundary. No CLAUSE boundary is marked if any of the following conditions is met.

1. The CCN is the first or second element of the sentence.
2. The WORD "," (element of PNT) precedes the CCN.
3. The WORDs "either" or "neither" occur proceeding the CCN elements "or" or "nor," respectively.

If none of these conditions is met, a CLAUSE boundary is marked.

2.1.2. Rules for the Class REL

If REL is preceded by PRP, then the PRP is marked as the CLAUSE boundary. If REL is preceded by CCN, the CCN is marked as the CLAUSE boundary. Otherwise, the REL is marked as the CLAUSE boundary.

2.1.3. Rules for the Class THT

In the assignment of WORDs to their respective classes by MYRA, THT is treated as a subset of the AMB class. CAP/I recognizes THT and makes several checks to determine whether THT is a CLAUSE marker. THT does not mark a CLAUSE if:

1. THT is the first element of a SENTENCE;
2. THT is preceded by PRP;
3. The CLAUSE in which THT is contained is initiated by PRP;
4. THT is followed by CCN;
5. THT is preceded by PNT.

2.1.4. Rule for the Class PNT

CAP/I identifies any PNT as a CLAUSE boundary unless the PNT is the first or second element of the SENTENCE.

Phase I of CAP/I terminates when

- 1) each occurrence of CCN, REL, THT and/or PNT in a SENTENCE has been examined;
- 2) the appropriate rules have been applied; and
- 3) the preliminary CLAUSE boundaries have been marked.

2.2. Phase Two of CAP/I

This phase concerns the application of rules which involve elements of the classes SCN, INF and PTC. When one of these elements is encountered in a SENTENCE, a new CLAUSE is indicated. The rules of phase 2 are described below.

2.2.1. Rule for the Class SCN

When an element of SCN is encountered in a SENTENCE, a new CLAUSE boundary is marked. The CLAUSE terminates when a new CLAUSE boundary is encountered or when EOS is reached.

2.2.2. Rule for the Class INF

When an element of INF is found, the immediately proceeding PRP is marked as a new CLAUSE boundary. The CLAUSE terminates as for SCN.

2.2.3. Rule for the Class PTC

A CLAUSE is marked whenever an element of PTC is found which is immediately preceded by PRP or SCN. If neither PRP nor SCN immediately precedes PTC, the CLAUSE boundary is marked at PTC.

2.3. Experimental Results from CAP/I

The procedures of CAP/I are applied to the NAMES and RELATIONS that have been identified by MYRA. As mentioned in Chapter III, MYRA was tested using three English texts comprising some 6,000 words. The unaltered output of MYRA was used to test CAP/I; hence CAP/I was tested on automatically classified NAMES and RELATIONS rather than on NAMES and RELATIONS manually classified. This was done to provide a measure of how MYRA and CAP/I operate together as a system. Sample output produced by CAP/I is given in Figure 4.1. Complete output for the article The Old Man and the Sea is given in Appendix C.

Like MYRA, CAP/I is programmed in PL/I for the IBM S/370-165 computer system. 126,000 bytes of main storage are required for the programs and for working storage. CAP/I operates at the rate of 15,000 words per minute.

2.3.1. Identification and Analysis of CAP/I Errors

The output of CAP/I was examined for errors using the same basic criteria as employed in the analysis of errors made by MYRA (Chapter III, Section 3.3). Errors were classified according to whether a CLAUSE was incorrectly identified or incorrectly delimited. For example, if, in the SENTENCE

The girl sitting on the stair won first prize.

the ADJECTIVAL CLAUSE "sitting on the stair" were not recognized, then one error would be noted for that fact, and one error would be recorded for the fact that the CLAUSE "The girl won first prize" would be

303. The old man had taught the boy

DTR	ADJ	NON	AUX	VRB	DTR	NON
-----	-----	-----	-----	-----	-----	-----

304. to fish

PRP	VRB
-----	-----

305. and the boy loved him.

CNJ	DTR	NON	VRB	PRN	EOS
-----	-----	-----	-----	-----	-----

Figure 4.1 Sample output produced by CAP/I from The Old Man
and the Sea.

incorrectly delimited.

The overall results of this error analysis for each of the texts tested are given in Table 4.1. An average accuracy of 62% was found for the delimitation of CLAUSES, and an average accuracy of 85% was determined for the identification of CLAUSES.

2.3.2. Conclusions From Analysis of CAP/I Results

In analyzing the results of CAP/I, the most significant problem was that of correctly recognizing a CLAUSE. Incorrect CLAUSE recognition also meant incorrect CLAUSE delimitation. Many of the errors committed by CAP/I could be corrected by incorporating rules which examined the patterns of NOMINAL, PRIMARY and SECONDARY PHRASES. Thus it will be suggested in Section 4 how an improved CLAUSE identification program might be developed to take advantage of such data. The necessary data are provided by the program to be described next.

3. Procedures for the Identification of PHRASES (PAP)

The procedures described in this section have been designed to identify and classify certain COMPOSITE and COMPLEX NAMES called, in general terms, PHRASES. Four types of PHRASE are identified by PAP:

NOMINAL PHRASE
PRIMARY PHRASE
SECONDARY PHRASE
ADV PHRASE

The first three PHRASE types have been defined in Chapter II, Section 3.4.3. An ADV PHRASE is defined as any occurrence of the class ADV outside the boundaries of a NOMINAL, PRIMARY or SECONDARY PHRASE.

Table 4.1 Accuracy of the Output of CAP/1. 28

NAME of Document	Correct Delimitation of CLAUSES %	Correct Identification of CLAUSES %
"...Precise Definition of 'Algorithm'"	82%	59%
<u>The Old Man and the Sea</u>	87%	76%
"The Clavichord..."	87%	50%

28. Based upon input produced by MYRA (see Table 3.7).

PAP operates in essentially two phases. In the first phase, certain of the WORD classes identified by MYRA that are linked by CCP are reduced to a single element of the appropriate class. Thus, a sequence such as NON CCP NON would be reduced to NON. In the second phase, PHRASEs are identified, delimited and classified.

3.1. Procedures for the Reduction of Certain COMPLEX NAMES to SIMPLE NAMES

PAP first examines a SENTENCE for CCPs which conjoin two SIMPLE NAMES or two SIMPLE RELATIONS. These triples, which are COMPLEX NAMES, are reduced by PAP to SIMPLE NAMES or RELATIONS. These new NAMES or RELATIONS retain the WORD class assignments made earlier by MYRA. The rules for effecting these reductions are as follows.

... NON CCP NON ...	⇒	... NON ...
... ADJ CCP ADJ ...	⇒	... ADJ ...
... ADV CCP ADV ...	⇒	... ADV ...
... VRB CCP VRB ...	⇒	... VRB ...
... PRP CCP PRP ...	⇒	... PRP ...
... AUX CCP AUX ...	⇒	... AUX ...

When the pattern PRN CCP PRN is found, the elements of PRN are examined for the following patterns.

... ('I' | 'she' | 'he' | 'we' | 'they') CCP ('me' | 'her' | 'him' | 'us' | 'them')
 ... ('me' | 'her' | 'him' | 'us' | 'them') CCP ('I' | 'she' | 'he' | 'we' | 'they')

If one of these patterns is encountered, no reduction is effected. Otherwise, the following rule is applied.

... PRN CCP PRN ... ⇒ ... PRN ...

3.2. Procedures for the Identification of PHRASEs

In the second phase of PAP, PHRASEs are identified, delimited and typed. The necessary procedures are based upon both the definition of each PHRASE type and upon the order in which NAMES and RELATIONS that may constitute the PHRASEs occur. The definitions of the four types of PHRASE are given in Table 4.2.

In this phase of PAP, the class of each WORD in the SENTENCE is examined in a left-to-right manner. The class of the first WORD in the SENTENCE determines the type of PHRASE to be delimited at that point. Consider, for example, the following SENTENCE and corresponding classes for each WORD.

The boy quickly ran down the trail.

DTR NON ADV VRB PRP DTR NON EOS

The class of the first WORD in the SENTENCE is DTR. This signals a NOMP (NOMINAL PHRASE) and calls for the application of pattern 1 and 2 of Table 4.2. Since the second element of SENTENCE is NON, either of the patterns is satisfied and the NOMP is isolated.

The next element in our example SENTENCE is ADV, which initiates the application of patterns 17, 18 and 21 (Table 4.2). Since the next SENTENCE element is VRB, pattern 17 is eliminated, as is pattern 21. Hence pattern 18 applies and a PRMP (PRIMARY PHRASE) is isolated.

The remainder of the SENTENCE is processed by PAP in a similar manner. The complete processing of the example SENTENCE by PAP would yield

Table 4.2 Rules for the Identification and Characterization of PHRASEs.

RULE	WORD CLASS PATTERN REQUIRED	PHRASE TYPE INDICATED	PHRASE LIMITS	
			INITIAL	FINAL
1.	$DTR^n ((ADV^n ADJ) ADJ^n) NON$	NOMP	/DTR	NON/
2.	$DTR^n ((INT ADJ) ADJ^n) NON$	NOMP	/DTR	NON/
3.	$ADJ^n NON$	NOMP	/ADJ	NON/
4.	NON	NOMP	/NON	NON/
5.	$INT ADJ^n NON$	NOMP	/INT	NON/
6.	PRN	NOMP	/PRN	PRN/
7.	$PRN_2 ADJ$	NOMP	/PRN ₂	ADJ/
8.	REL NOMP	NOMP	/REL	NOMP/
9.	$NEV AUX^n ((INT) ADV) VRB$	PRMP	/NEV	VRB/
10.	$NEV AUX^n$	PRMP	/NEV	AUX/
11.	$NEV PTC^n$	PRMP	/NEV	PTC/
12.	$NEV 'to' AUX ((INT) ADV) VRB$	PRMP	/NEV	VRB/
13.	$NEV 'to' ((INT) ADV) VRB$	PRMP	/NEV	VRB/
14.	$AUX^n NEG VRB$	PRMP	/AUX	VRB/
15.	$AUX^n VRB NEG$	PRMP	/AUX	NEG/
16.	$AUX^n ((INT) ADV) VRB$	PRMP	/AUX	ADV/
17.	$AUX^n VRB ((INT) ADV$	PRMP	/AUX	ADV/
18.	ADV AUX VRB	PRMP	/ADV	VRB/

Table 4.2 (continued)

RULE	WORD CLASS PATTERN REQUIRED	PHRASE TYPE INDICATED	PHRASE LIMITS	
			INITIAL	FINAL
19.	ADV VRB	PRMP	/ADV	VRB/
20.	VRB ADV	PRMP	/VRB	ADV/
21.	PRP ⁿ NOMP	SECP	/PRP	NON/
22.	ADV	ADVP	/ADV	ADV/
23.	PRP (PNT)	ADVP	/PRP	PRP/
24.	PRP (EOS)	ADVP	/PRP	PRP/

The boy quickly ran down the trail.

DTR	NON	ADV	VRB	PRP	DTR	NON	EOS
	NOMP		PRMP		SECP		

Processing is terminated when EOS is encountered.

3.3. Procedures for the Reduction of Certain COMPLEX NAMES to PHRASES

Once PHRASEs have been delimited, it may be found that two or more of them are conjoined by CCP. For example,

DTR	NON	ADV	VRB	PRP	DTR	NON	CCP	PRP	ADJ	NON	EOS
	NOMP		PRMP		SECP				SECP		

Under these circumstances, it may be desirable to reduce such a COMPLEX NAME to a PHRASE of the same type as those in the COMPLEX NAME. Such a reduction is accomplished in the same manner as reduction of COMPLEX NAMES and RELATIONS to SIMPLE NAMES and RELATIONS (Section 3.1). The following rules serve for the reduction of COMPLEX NAMES to PHRASEs.

... ADVP CCP ADVP ...	⇒	... ADVP ...
... NOMP CCP NOMP ...	⇒	... NOMP ...
... PRMP CCP PRMP ...	⇒	... PRMP ...
... SECP CCP SECP ...	⇒	... SECP ...

Thus, in the following example, two NOMINAL PHRASEs are linked by CCP are

The suitcases and packages were left on the plane.

DTR	NON	CCP	NON	AUX	VRB	PRP	DTR	NON	EOS
	NOMP		NOMP		PRMP		SECP		

reduced to a single NOMP, hence:

DTR	NON	CCP	NON	AUX	VRB	PRP	DTR	NON	EOS
		NOMP			PRMP		SECP		

3.3.1. Reductions Not Involving CCP

ADJECTIVAL SECONDARY RELATIONS introduce PHRASEs that may be combined with a preceding NOMINAL PHRASE to form a new NOMINAL PHRASE. To accomplish such a reduction, it is first necessary to distinguish between ADJECTIVAL and ADVERBIAL SECONDARY RELATIONS. The distinction between these two classes is based upon the structure in which the RELATION occurs. The following rules serve to distinguish ADJECTIVAL SECONDARY RELATIONS from ADVERBIAL SECONDARY RELATIONS which are therefore defined by default.

1. The SECONDARY RELATION "of" introduces an ADJECTIVAL SECONDARY PHRASE if it initiates a SECONDARY PHRASE and if it is preceded by a NOMINAL PHRASE.
2. Any SECONDARY RELATION introduces an ADJECTIVAL SECONDARY PHRASE if it initiates a SECONDARY PHRASE that is preceded by a NOMINAL PHRASE and followed by a PRIMARY PHRASE.

An ADJECTIVAL SECONDARY RELATION is considered to be an attribute of the NOMINAL PHRASE which precedes it. Thus, the NOMINAL PHRASE and the SECONDARY PHRASE are reduced to a single NOMINAL PHRASE. As the following example illustrates, the SENTENCE

The present for the children was filled with bags of candy.

DTR	NON	PRP	DTR	NON	AUX	VRB	PRP	NON	PRP	NON	EOS
	NOMP		SECP		PRMP		SECP		SECP		

becomes²⁹

-
29. By using such procedures, the SENTENCE "The children's present was filled with candy bags." would be assigned the same PHRASE structure.

The present for the children was filled with bags of candy.

DTR	NON	PRP	DTR	NON	AUX	VRB	PRP	NON	PRP	NON	EOS
		NOMP				PRMP			SECP		

3.4. Results of Tests

The WORD and CLAUSE assignments produced by MYRA and CAP/I form the input to PAP. PAP was tested on the data generated from the three articles mentioned earlier. An average accuracy of 90% was attained in correctly delimiting PHRASEs and an accuracy of 92% was attained in correctly identifying PHRASEs. The results produced by PAP for each of the three documents analyzed are presented in Table 4.3. Sample output produced by PAP is illustrated in Figure 4.2. More extensive output may be found in Appendix C.

Most of the error which occurred both in delimiting and identifying PHRASEs was caused by errors in MYRA and CAP. Only a small percentage of error was directly caused by the definition of the structure of PHRASE types. A detailed analysis of the results generated by PAP is presented in Appendix E.

In this and the preceding chapter, three programs have been described: MYRA, CAP/I and PAP. The organization of these programs into a system for language analysis is illustrated in Figure 4.3. As alluded to in the discussion of CAP/I, this organization is somewhat illogical in that knowledge of PHRASE boundaries would be of help in identifying and delimiting CLAUSES, whereas the converse does not hold. Therefore, this organization has been modified as depicted in Figure 4.4. PAP now

Table 4.3 Accuracy of the Output of PAP.³⁰

NAME of Document	Correct Delimitation of PHRASES %	Correct Identification of PHRASE Type %
"...Precise Definition of 'Algorithm'"	90%	92%
<u>The Old Man and the Sea</u>	93%	94%
"The Clavichord..."	87%	90%

30. Based upon input produced by CAP/I (see Table 4.1).

303. The old man had taught the boy

DTR	ADJ	NON	AUX	VRB	DTR	NON
	NOMP			PRMP		NOMP

304. to fish

PRP	VRB
	PRMP

305. and the boy loved him.

CNJ	DTR	NON	VRB	PRN	EOS
	NOMP		PRMP	NOMP	

Figure 4.2 Sample output produced by PAP from The Old Man
and the Sea.

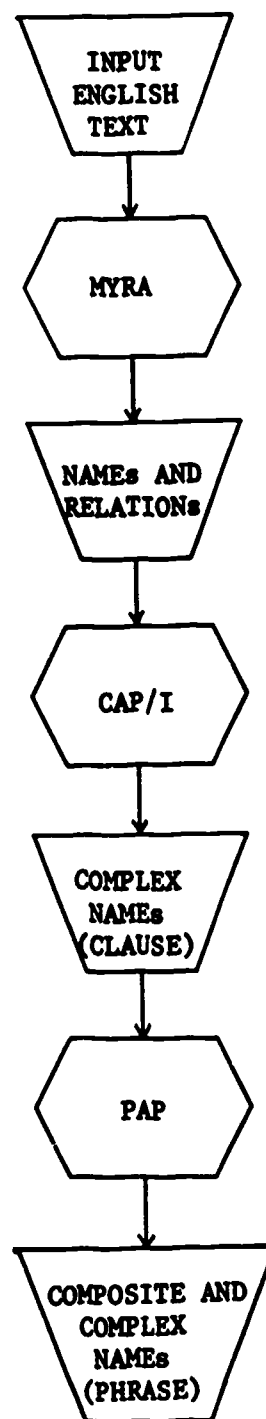


Figure 4.3 The operation of the language analysis procedures.

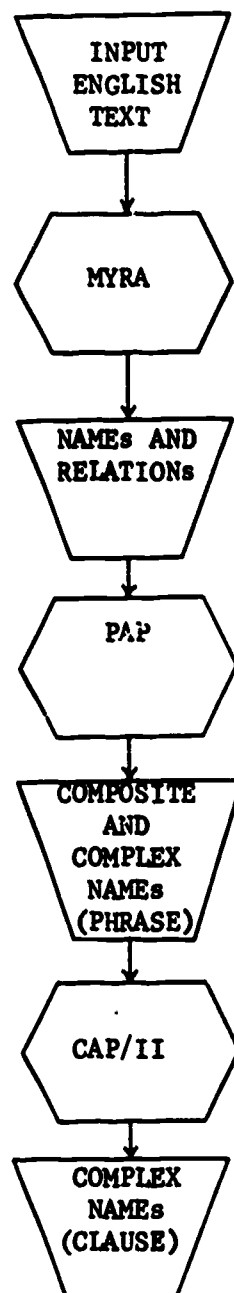


Figure 4.4 The operation of the language analysis procedures with the refined version of clause recognition.

derives its input directly from MYRA and a new CLAUSE program, CAP/II, has been designed which accepts input both from MYRA and from PAP. CAP/II is described in the next section.

4. Improved Procedures for the Identification of CLAUSES

CAP/II incorporates not only those procedures embodied in CAP/I, but also contains a number of additional procedures that take advantage of the output of PAP. Phase I of CAP/II corresponds with phase I of CAP/I. Phase II of CAP/II is carried out in the same way as for CAP/I, but now no tests are made involving the classes INF and PTC. After insertion of the CLAUSE markers, the strings are examined to determine the accuracy of the CLAUSE-boundary assignment. If a PRIMARY RELATION is not found within CLAUSE boundaries, one or both boundaries are deleted. If a CLAUSE initiates a SENTENCE, its right boundary is deleted; otherwise a right boundary is deleted until a PRIMARY PHRASE is found or until EOS. If no PRIMARY PHRASE is found, left boundaries are deleted, from right to left until a PRIMARY PHRASE is found. In the next step, CAP/II examines PHRASE patterns to determine whether CLAUSES are present which have not yet been identified.

4.1. Rules for Identifying CLAUSES which Involve the Class INF

CAP/II contains a number of rules for the identification of CLAUSES based upon patterns of PHRASEs. Five of these rules involve the Class INF (a class of PRIMARY RELATION). These rules are:

1. ... NOMP PRMP' ... \Rightarrow .../NOMP PRMP' ...

- *2. ... PRMP' SECPⁿ PRMP ... = .../PRMP' SECPⁿ/PRMP ...
- *3. ... PRMP' NOMP SECPⁿ PRMP ... = .../PRMP' NOMP SECPⁿ/PRMP ...
- *4. ... PRMP' NOMP SECPⁿ NOMP PRMP ... =
.../PRMP' NOMP SECPⁿ NOMP/PRMP ...
- *5. ... PRMP' SECPⁿ NOMP ... PRMP ... =
.../PRMP' SECPⁿ/NOMP ... PRMP ...

(where PRMP' is a PRIMARY PHRASE containing INF, $n \geq 1$, "/" marks a CLAUSE boundary and where rules marked with an asterisk are interpreted as placing a clause marker before PRMP' only if rule 1 cannot be applied).

If a PRMP' is immediately preceded by NOMP, the NOMP initiates the CLAUSE. Otherwise, PRMP' initiates the CLAUSE. The right boundary of each CLAUSE identified by rules 2-5 is marked as indicated in the rules.

4.2. Rules for Identifying CLAUSES which Involve the Class PTC

The rules in which PTC is involved are:

6. ... PRMP* SECPⁿ PRMP ... = .../PRMP* SECPⁿ /PRMP ...
7. ... PRMP* NOMP SECPⁿ PRMP ... =
.../PRMP* NOMP SECPⁿ /PRMP ...
8. ... PRMP* (NOMP SECPⁿ)² PRMP ... =
... /PRMP* (NOMP SECPⁿ)² /PRMP ...
9. ... PRMP* (NOMP SECPⁿ)² NOMP ... PRMP ... =
.../PRMP* (NOMP SECPⁿ)² /NOMP ... PRMP ...

(where PRMP* is a PRIMARY PHRASE containing PTC, $n \geq 0$ and "/" marks a CLAUSE boundary). PRMP* always initiates a CLAUSE. If PRMP* preceded by a NOMINAL PHRASE, the CLAUSE initiated by PRMP* is ADJECTIVAL (see Figure 2.7, Chapter II), otherwise the CLAUSE is NOMINAL.

4.3. Other CLAUSE Identification Rules

After all of the CLAUSE identification procedures so far described have been applied by CAP/II, a final check is made of the SENTENCE for any unidentified CLAUSES according to the following rules

10. ... PRMP SECPⁿ PRMP ... =
... PRMP SECPⁿ/PRMP ...
11. ... PRMP SECPⁿ NOMP PRMP ... =
... PRMP SECPⁿ/NOMP PRMP ...
12. ... PRMP NOMP SECPⁿ PRMP ... =
... PRMP/NOMP SECPⁿ PRMP ...

(where $n \geq 0$ and "/" marks a boundary between CLAUSES). These rules serve only to mark the boundary between CLAUSES. The remaining CLAUSE boundaries are determined by the other procedures in CAP/II (for instance, EOS would mark the right boundary of a CLAUSE whose left boundary was marked by use of one of the rules 10-12).

Although CAP/II has not been implemented, the procedures which it embodies have been tested manually and have been shown to give much better results than CAP/I. Implementation and testing of CAP/II are currently being carried out.

This chapter has described two programs which are currently in operation (CAP/I and PAP) and has described the design of a third program (CAP/II) which is currently under construction. These programs provide a means for analyzing English text in terms of NAMES and RELATIONS on several different levels (i.e., SIMPLE NAMES and RELATIONS, CLAUSES and PHRASES.) The output generated by these programs has been illustrated and the accuracy of the result has been detailed. A final

program developed in this research, and described in the next chapter, builds upon the results of MYRA, CAP/I and PAP.

References

1. A. Oettinger and S. Kur , , "Multiple-path Syntactic Analyzer," Proceedings of the International Federation of Information Processing Congress, Munich, 1962.
2. A. M. Zwicky, J. Friedman, B. C. Hall and D. E. Walker, "The MITRE Syntactic Analysis Procedure for Transformational Grammars ,"
Proceedings of the Fall Joint Computer Conference 27 (01), 1965, 317-326.
3. N. Sager and M. Salkoff, "Syntactic Analysis of Natural Language ,"
Advances in Computers 08 (00) 1969, 153-188.
4. W. A. Woods, "Transition Network Grammars for Natural Language Analysis," Communications of the ACM 13 (10), 591-606 (1970).
5. T. Winograd, "Understanding Natural Language," Cognitive Psychology 3, 1-191 (1972).
6. Woods, op. cit.
7. H. W. Buttleman, private communication, The Ohio State University, 1973.
8. W. Stolz, P. Tannenbaum and P. Carstensen, "A Stochastic Approach to the Grammatical Coding of English," Communication of the ACM 8 (6), 399-405 (1965).

CHAPTER V. THE FUNCTION OF NAMES AND RELATIONS.

1. Introduction

In previous chapters, procedures have been described which identify NAMES and RELATIONS on several levels of complexity. These procedures are largely structurally based. We turn now to a consideration of procedures which deal with the function of NAMES and RELATIONS within a SENTENCE. Function is here described in terms of the theory of case grammar formalized by Fillmore (1, 2).

A brief review of the notions of case (in the Fillmorian sense) is presented first, followed by a description of Fillmore's original theory. Then a modification of his theory developed for this research is presented and finally, procedures are described which categorize elements of a SENTENCE according to their function within this framework.

2. Case Grammar

The notions of case have been discussed for many years, but commonly in terms that modern linguists find to be of little practical utility. The traditional cases included nominative, genative, dative, and accusative. In contrast, case grammar is concerned with the role played by various elements of a sentence. Thus, typical cases are agent, object, experiencer, beneficiary, locative and so on.

Several intuitive definitions of this type of case role are to be found in the information storage and retrieval literature (3, 4, 5, 6). Case roles are also to be found in one guise or another in the work of Shank and Tesler (7), Winograd (8), Woods (9), Quillian (10) and others. (Some details of the work of these authors have been given in previous chapters.)

Case grammar was first presented by Charles Fillmore in 1966 (1). However, the classic statement of case grammar is found in "The Case for Case" presented by Fillmore in 1968 (2). Fillmore himself has modified his model of case grammar over the past few years (11, 12) and many workers involved in semantic analysis have adopted some form of case grammar. For a detailed description of Fillmore's positions and of work in which modifications of case grammar have been made for specific uses, the reader is referred to (12, 13). The model of case grammar given here differs from those presented by Fillmore in two major respects: first the PRIMARY RELATION is viewed as central and demands certain case roles; second, case roles must depend minimally upon extralinguistic evidence for their identification. The model of case grammar presented below builds upon Fillmore's work, and includes conceptualizations of Chafe and Cook (14, 15, 16, 17, 18).

2.1. The Basic Components of Case Grammar

Case grammar postulates that 1) the basic unit of text for analysis is the CLAUSE; 2) a CLAUSE consists of a series of NOMINAL and SECONDARY PHRASEs which are non-linearly related to a PRIMARY PHRASE. PRIMARY PHRASEs are partitioned into several categories, and NOMINAL

SECONDARY PHRASEs are partitioned into categories under the control of the category of PRIMARY PHRASE. A category of NOMINAL or SECONDARY PHRASE is called a case grammar role, or case role. The structure of a CLAUSE is represented in terms of the category of the PRIMARY PHRASE and in terms of the case roles assigned to the NOMINAL or SECONDARY PHRASEs. This representation is referred to as a case frame.

2.2. Categories of PRIMARY RELATION

Case grammar defines five categories of PRIMARY PHRASE:

AGENTIVE

BENEFACTIVE

EXPERIENCER

REFLEXIVE

STATIVE

The distinction between these categories is made on the basis of the identity of the MAIN PRIMARY RELATION in the PRIMARY PHRASE. For the BENEFACTIVE category, the MAIN PRIMARY RELATION elements are:

had
has
have
having

The EXPERIENCER category is defined by the following elements of MAIN PRIMARY RELATION.

died	love
doubt	remember
fear	see
feel	smell
hear	understand
hope	want
know	wish
like	wonder

The MAIN PRIMARY RELATIONS in the category STATIVE are:

am	being
are	is
be	was
been	were

Thus, the categories BENEFACTIVE, EXPERIENCER and STATIVE are defined ostensively. The category RELFEXIVE is defined as being comprised of those PRIMARY PHRASEs (other than those in the categories BENEFACTIVE, EXPERIENCER or STATIVE) which are preceded in a CLAUSE by exactly one NOMINAL PHRASE and which are followed by no NOMINAL PHRASE. The category AGENTIVE is defined by default.

2.3. Case Roles

Two classes of case role are identified: essential (also called nuclear, propositional or major) and peripheral (also called modal or minor). In an intuitive sense, essential cases are assigned to those elements of a CLAUSE which can be interpreted as answering the questions "who," "what," "which"; peripheral cases are assigned to elements which answer questions like "when," "how," "why," "where," "for what." In general, essential cases are those demanded by a PRIMARY PHRASE, while peripheral cases are usually, if not always, optional.

The case roles and their definitions, according to Fillmore (19), are given in Table 5.1. As pointed out above, these case roles have been modified somewhat for the purposes of this research. The main purpose of the changes was to enable the identification of case roles by algorithmic means. The case roles used in this work, together with their definitions are presented in Table 5.2. The case roles presented

Table 5.1 Case Roles and their Definitions According to Fillmore (19).

TYPE	CASE ROLE	SYMBOL	DEFINITION
ESSENTIAL	AGENTIVE	A	instigator of the action, animate
	EXPERIENCER	E	affected by the action, animate
	INSTRUMENTAL	I	force or object causing action of state
	OBJECTIVE	O	semantically most neutral case
	SOURCE	S	the origin or starting point
	GOAL	G	the object or end point
PERIPHERAL	LOCATIVE	L	spatial orientation of the action
	TIME	T	temporal orientation of the action
	COMITATIVE	C	accompaniment role, animate
	BENEFACTIVE	B	benefactive role, animate

Table 5.2 Definition of Essential and Peripheral Case Roles as Used in the Present Research.

CASE ROLE	SYMBOL	DEFINITION OF CASE ROLE
AGENT	A	the source of the action specified by the PRIMARY PHRASE.
EXPERIENCER	E	the one who experiences the feeling, sensation, etc., specified by the PRIMARY PHRASE.
BENEFICIARY	B	the possessor (in its broadest sense) of some thing, whether the possession be temporary or permanent, positive or negative.
OBJECTIVE	O	the receiver of the action described by the PRIMARY PHRASE.
LOCATIVE	L	the place where the action described by the PRIMARY PHRASE occurs.
TIME	T	the time when the action described by the PRIMARY PHRASE occurs.
MANNER	M	the way in which the action described by the PRIMARY PHRASE is performed.
COMITATIVE	C	the accompaniment case, a subject accompanying the source of the action described by the PRIMARY PHRASE.
CAUSE	Cs	the case giving the reason for the action described by the PRIMARY PHRASE.
PURPOSE	P	the case giving the purpose of the action described by the PRIMARY PHRASE.

by Fillmore (10) are contrasted with those used in the present work in Table 5.3. To give the reader some notion of the applications of case grammar, Table 5.4 presents English sentences illustrative of each of the five categories of PRIMARY PHRASE, together with the case-role assignments for the various elements of each sentence.

3. Identification of Case Roles

In this research, case roles are assigned to NOMINAL PHRASEs, SECONDARY PHRASEs and CLAUSES (a departure from Fillmore's procedures). The conditions under which such assignments are made, as well as the nature of the assignments, will be described in this section. In general, case grammar analysis is carried out on the CLAUSE.

3.1. Essential Case Roles

The assignment of essential case roles within the CLAUSE is determined by the following rules.

1. For the AGENTIVE category: A case preceding and E and O cases following the PRIMARY PHRASE in that order, except if the PRIMARY PHRASE is passive,³¹ O case preceding and E and A cases following the PRIMARY PHRASE in that order.
2. For the BENEFACTIVE category: B case preceding and O case following the PRIMARY PHRASE, except if the PRIMARY PHRASE is passive³¹ in which case these assignments are reversed.
3. For the EXPERIENCER category: E case preceding and O case following the PRIMARY PHRASE, except if the PRIMARY PHRASE is passive³¹ in which case these assignments are reversed.
4. For the REFLEXIVE category: A - O case preceding the PRIMARY PHRASE.

31. A passive PRIMARY PHRASE is defined as follows. If the PRIMARY PHRASE is initiated by AUX and if it is followed by a SECONDARY PHRASE introduced by "by" in the same CLAUSE, the PRIMARY PHRASE is passive.

Table 5.3 Case Roles as After Fillmore and as Used in this Research.

TYPE	CASE ROLES		SYMBOL
	FILLMORE	PRESENT RESEARCH	
ESSENTIAL	AGENT	AGENT	A
	EXPERIENCER	BENEFICIARY	B
	GOAL	EXPERIENCER	E
	INSTRUMENTAL	OBJECT	O
	LOCATIVE		
	OBJECT		
	SOURCE		
PERIPHERAL	BENEFACTIVE	CAUSE	Cs
	COMITATIVE	COMITATIVE	C
	LOCATIVE	LOCATIVE	L
	TIME	MANNER	M
		PURPOSE	P
		TIME	T

Table 5.4 Case-Role Assignments for Sample CLAUSES each
Containing one of the Categories of PRIMARY PHRASE.

PRIMARY PHRASE CATEGORY	ILLUSTRATIVE EXAMPLES
STATIVE	The boy is a man today. O stative O T
BENEFACTIVE	I have the book in the library. B benefactive O L
EXPERIENCER	The little girl liked ice cream E experiencer O with chocolate syrup. M
REFLEXIVE	The bird flew A-O reflexive
AGENTIVE	The cyclist hit the car A agentive O He gave me the letter A agentive E O

5. For the STATIVE category: 0 case preceding and/or following the PRIMARY PHRASE.

The case frames to which these rules give rise are summarized in Table 5.5. The way in which these rules are applied is described in Section 4.

3.2. Peripheral Case Roles

In general, the assignment of peripheral cases is made to SECONDARY PHRASEs and is controlled by the SECONDARY RELATION which introduces the PHRASE. The time case presents an exception to this general statement. WORDs that signal the time case must be defined ostensively. They are:

again	first	once	week
already	frequently	still	weeks
always	later	then	when
day	month	time	year
days	months	today	years
early	never	tomorrow	yet
finally	now	tonight	

If a SECONDARY PHRASE contains one of these WORDs it is assigned the T case. Otherwise, the following rules apply.

1. A SECONDARY PHRASE initiated by "to" is assigned L case.
2. A SECONDARY PHRASE initiated by "by" is assigned M case.
3. A SECONDARY PHRASE initiated by "for" is assigned P case.
4. A SECONDARY PHRASE initiated by "with" or by "without" is assigned C case if the PHRASE contains the NAME of an animate entity,³² and the M case otherwise.
5. A SECONDARY PHRASE initiated by any other SECONDARY RELATION is assigned L case.

32. An animate entity is defined to be any PHRASE containing one of the WORDs he, she, her, him, they, them, we or us.

Table 5.5 Case Frames for Essential Cases as Prescribed by the PRIMARY PHRASE Category.

CATEGORY	PASSIVE ³² ?	CASE FRAMES
AGENTIVE	No Yes	A AGENTIVE E, O O AGENTIVE E, A
BENEFACTIVE	No Yes	B BENEFACTIVE O O BENEFACTIVE B
EXPERIENCER	No Yes	E EXPERIENCER O O EXPERIENCER E
REFLEXIVE	---	A-O REFLEXIVE
STATIVE	---	O STATIVE O STATIVE O

32. A passive PRIMARY PHRASE is defined as follows. If the PRIMARY PHRASE is initiated by AUX and if it is followed by a SECONDARY PHRASE introduced by "by" in the same CLAUSE, the PRIMARY PHRASE is passive.

The way in which these rules are applied is described in Section 4.

3.3. Assignment of Case Roles to CLAUSES

In addition to the case-role assignments described above, work has begun to define the case roles of CLAUSES. This is a major extension of the original Fillmorean conception of case, one that follows logically from the NAME/RELATION view of language espoused in this dissertation. The rules which have so far been developed for the assignment of case roles to CLAUSES are restricted to CLAUSES which contain an element of PTC and which are initiated by a SECONDARY RELATION.

The rules are:

1. A CLAUSE initiated by "from" receives the Cs case (e.g., His hands were tough from handling heavy cords.).
2. A CLAUSE initiated for "for" receives the P case (e.g., The equipment was used for testing compounds.).
3. A CLAUSE initiated by "by" receives the M case (e.g., He tested the compounds by using a new method.).
4. A CLAUSE initiated by "with" receives the O case (e.g., She was finished with typing the dissertation.).
5. A CLAUSE initiated by "of" and preceded by the sequence VRB ADJ receives the O case (e.g., I am sick of making cakes.).
6. A CLAUSE initiated by "of" and preceded by a NOMINAL PHRASE is adjectival and receives the same case as the preceding NOMINAL PHRASE (e.g., The process of making clay.).

4. The Case Grammar Program, CGP

The definitions and rules presented in the preceding section have been incorporated into a program called CGP. The program is written in PL/I for the IBM 370-165 computer system and requires 126,000 bytes of main storage. Execution times are in the range 15,000 words per

minute. CGP accepts as input the output of MYRA, PAP and CAP/I. It operates in two phases. Phase I is concerned with the categorization of the PRIMARY PHRASE within the CLAUSE. STATIVE, BENEFACTIVE and EXPERIENCER categories are determined by dictionary look-up. The REFLEXIVE category is determined by considering the number of NOMINAL PHRASEs surrounding the PRIMARY PHRASE. If only one is found and it precedes the PRIMARY PHRASE, the category REFLEXIVE is identified. All other PRIMARY PHRASEs are AGENTIVE.

After the PRIMARY PHRASEs have been categorized, the essential cases are assigned. Such assignments generally follow the rules of Section 3.1. In addition, if SECONDARY PHRASEs containing NAMES indicative of animate objects are introduced by the SECONDARY RELATIONS "to" or "for", the SECONDARY PHRASE is assigned the E or B case, respectively.

Peripheral cases are assigned to SECONDARY PHRASEs based upon the rules given in Section 3.2. Case assignments to CLAUSES follow the rules of Section 3.3.

A flow diagram of CGP is given in Figure 5.1, and typical results produced by the program are illustrated in Figure 5.2. More extensive output will be found in Appendix C.

5. Experimental Testing of CGP

The case grammar program CGP was tested using the same documents as processed by MYRA, CAP/I and PAP. The output from these three programs formed the input to CGP. The results were analyzed by comparing

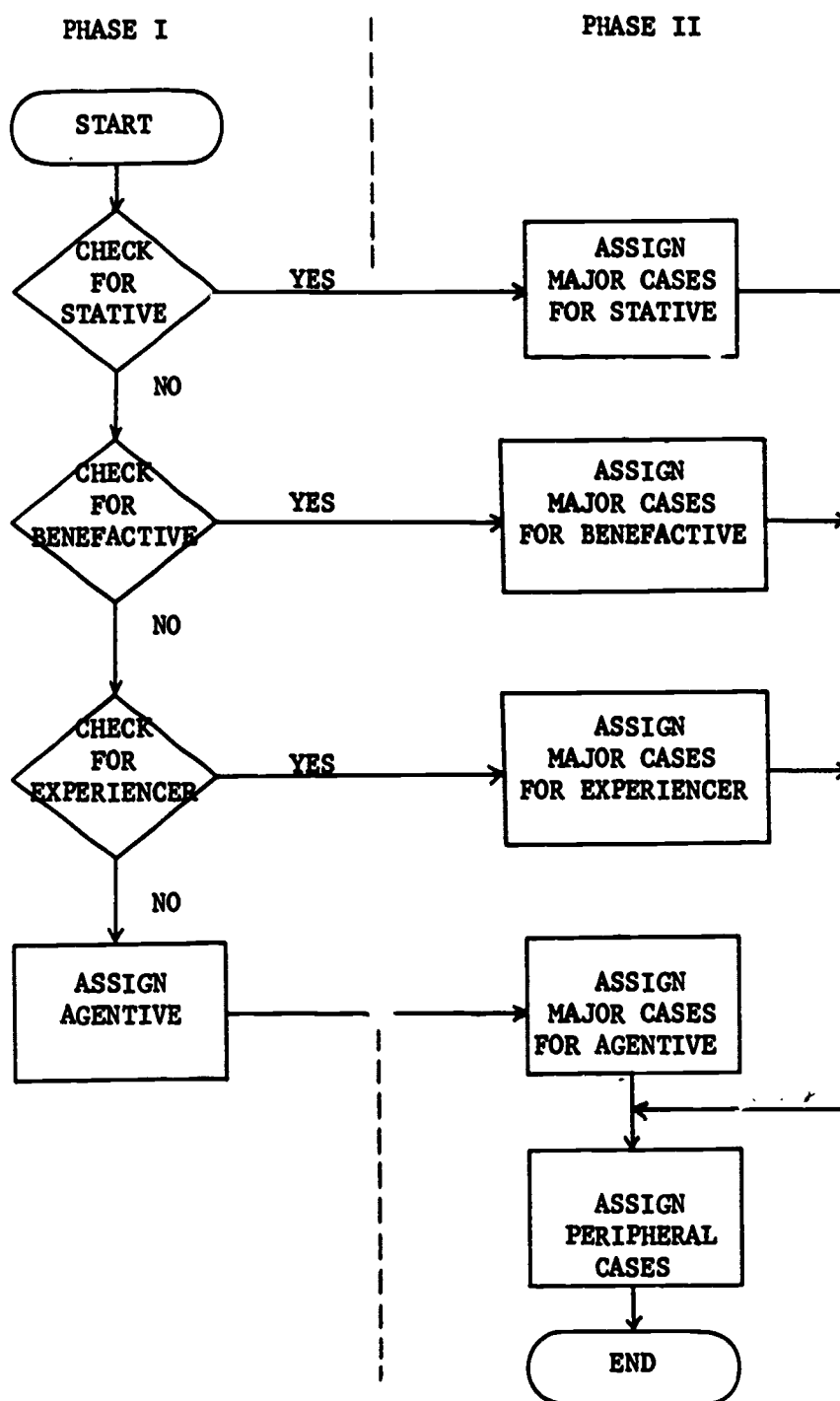


Figure 5.1 A flow diagram of CGP.

303. The old man had taught the boy

DTR	ADJ	NON	AUX	VRB	DTR	NON
	NOMP		PRMP		NOMP	
	AGENT				OBJECT	

304. to fish

PRP	VRB
PRMP	

305. and the boy loved him .

CNJ	DTR	NON	VRB	PRN	EOS
	PRMP		PRMP		
	AGENT				

Figure 5.2 Sample output produced by CGP from The Old Man and the Sea.

the cases assigned with the cases I would intuitively expect for the construction. Differences between assignment and expectation were treated as errors committed by CGP. The results of this error analysis are given in Table 5.6. PAP and CGR operate in 126,000 bytes of storage at the rate of 7,700 words per minute.

6. Summary

The program described in this chapter represents the first implementation of case grammar. While the accuracy achieved was only in the range of 75%, many of the errors were due not to CGP, but to the programs which produced the input to the case grammar program. Given accurate input, CGP may reasonably be expected (based upon preliminary studies) to achieve greater than 95% accuracy. Further work to refine the case grammar program are currently under way.

Table 5.6 Accuracy Attained in Case Grammar Assignment Made by CGP.

DOCUMENT	ACCURACY ATTAINED
"..Precise Definition of 'Algorithm'"	70%
<u>The Old Man and the Sea</u>	83%
"The Clavichord and How to Play It"	65%

References

1. C. J. Fillmore, "A Proposal Concerning English Prepositions," Georgetown University Round Table on Languages and Linguistics, 19th Monograph, 1966.
2. C. J. Fillmore, "The Case for Case," in Universals in Linguistic Theory, Bach, et al. (ed.), Holt, Rinehart and Winston, Inc., New York, 1968.
3. B. A. Montague, "Patent Indexing by Concept Coordination Using Links and Roles," American Documentation 13, 104-11 (1962).
4. G. Salton, "Manipulation of Trees in Information Retrieval," Communications of the ACM, 05 (02), 103-144 (1962).
5. H. Skolnik, "The Multiterm Index: A New Concept in Information Storage and Retrieval," Journal of Chemical Documentation 10 (02), 81-84 (1970).
6. R. E. Maizell, J. F. Smith and T. E. Singer, Abstracting Scientific and Technical Literature, Wiley-Interscience, New York, 1971, p. 74.
7. Shank and Tesler, "Conceptual Dependency Parser," Computational Linguistics, Coling, preprint No. 2, Fock, Stockholm.
8. T. Winograd, "Understanding Natural Language," Cognitive Psychology 3, 1-191 (1972).
9. W. A. Woods, "Transition Network Grammars for Natural Language Analysis," Communication of the ACM 13 (10), 591-606 (1970).
10. M. R. Quillian, "The Teachable Language Comprehender: A Simulation Program and Theory of Language," Communication of the ACM 12 (8), 459-476 (1969).
11. C. J. Fillmore, "Improvement in Case Grammar, 1970," presented at the Summer Institute of the Linguistic Society of America at The Ohio State University, 1970.
12. C. J. Fillmore, "Some Problems for Case Grammar," Working Papers in Linguistics, No. 10, The Ohio State University, 1971.

13. W. A. Cook, S.J., "Case Grammar, 1972," preprint for Languages and Linguistics Working Papers, Georgetown University Press, Washington, D.C.
14. W. L. Chafe, Meaning and the Structure of Language, The University of Chicago Press, 1970.
15. W. A. Cook, S.J., "Case Grammar: From Roles to Rules", Languages and Linguistics Working Papers No. 1, Georgetown University Press, Washington, D.C., 1970, pp. 14-30.
16. W. A. Cook, S.J., "Case Grammar as a Deep Structure in Tagmemic Analysis," Languages and Linguistics Working Papers No. 2, Georgetown University Press, Washington, D.C., 1971, pp. 1-9.
17. W. A. Cook, S.J., "Improvements in Case Grammar 1970", Languages and Linguistics Working Papers No. 2, Georgetown University Press, Washington, D.C., 1971, pp. 10-23.
18. W. A. Cook, S.J., "A Case Grammar Matrix," Languages and Linguistics Working Papers, No. 6, Georgetown University Press, Washington, D.C., 1972, pp. 15-48.
19. C. J. Fillmore, 1970, op.cit.

CHAPTER VI. LANGUAGE ANALYSIS AND INDEXING

By definition, the function of a metalanguage in semantic analysis is to singularize or to differentiate--the two notions are interchangeable--the documents of a corpus by the interplay of complex correspondances between natural language formulations and equivalent expressions in a metalanguage.

J. C. Gardin, Semantic Analysis Procedures in the Sciences of Man

1. Introduction

In previous chapters of this dissertation a series of language analysis procedures has been described. These procedures are largely syntax based; the strongest appeal to interpretational characteristics is made in the case grammar analysis procedures of Chapter V. In the present chapter, procedures are outlined which permit the construction of structural representations of English sentences (the metalanguage of Gardin (1)). Since the basic purpose for all the work described in this document is the development of means of producing improved indexes, a brief statement of how such structural representations relate to indexing is in order. Following that, a review of some closely related work is given. The remainder of the chapter is then devoted to a discussion of the proposed structural representations.

2. Indexes and Indexing

It is conventional to mention the exponential growth of the literature of various fields as the main cause of the difficulty one

meets in trying to find data pertinent to his needs. The quantity of documents makes researching the literature difficult, but this difficulty is not the central problem in the array of problems which the information storage and retrieval specialist hopes to solve. In fact, a statement made some 300 years ago by Howell is still adequate to sum up the important problem which must be solved if published data are ever to be readily accessible when one needs the data:

The reason why there is no Table or Index added hereunto is, that every page is so full of signal remarks that were they couched in an Index it would make a volume as big as the book and so make the Postern Gate to bear no proportion to the building. (2)

Much work has been done to overcome the problem which Howell solved by deciding not to include an index to his book at all. Most of this work has been directed toward increasing the amount of data present in indexes, without a concomitant increase in physical size of the index. If one can speak of "data density," these efforts have all been directed toward increasing the "data density" of the index. One of the very popular techniques has been that of keyword indexing. The central assumption of keyword indexing is that more complex concepts (data) can be formed at search time through appropriate application of Boolean algebraic functions (relations) to combinations of keywords. The inadequacy (3) of such an approach is due to the fact that most of the relationships which humans impute between data elements cannot be expressed in terms of any combination of Boolean functions. As a consequence of the failure of keyword indexing to yield satisfactory indexes, attempts to improve keyword indexing, to expand unit term

indexing to multi-term indexing (4), to improve keyword-in-context indexes (5), to produce articulated indexes (6), to simulate human memory (7), and to produce general representations of language (8) have been made which are all concerned with retention (and explication) of relations among data elements.

It might be reasonable to suppose that a document is its own best representation as Howell concluded. But this supposition rests upon the further assumption that the words in the document have the same significance for the reader as for the author. Thus, two important factors must be accounted for in indexing any document. These factors are:

- 1) The structural properties of a document. (A document conveys data to the reader through the unique organization of language elements within the document.)
- 2) The interpretational characteristics of a document. (A document conveys data to the reader through the significance imputed to these language elements outside the framework of that document.)

Much attention has been paid the latter factor. Thesauri, authority lists, dictionaries and other devices have been proposed, discussed, designed, built, used and scrapped by many workers. At the same time, the organization and structure of the document has been treated (when recognized at all) as something to be got over as soon as possible. As a result of this attitude, documents are deliberately shorn of the relations they contain at the indexing stage.

To improve indexes, indexing methods must be developed which

- a) retain as much of the data contained in a document as possible;
- b) retain and make explicit relationships among elements of these data;
- c) add data not expressed in a document but which derive from relations among elements of data in a document and elements of data contained in analysis documents (such as thesauri and dictionaries);
- e) develop efficient methods for deriving the index material from text, with scant appeal to the "meaning" of text data.

This research has been directed toward the development of language analysis procedures through which one might achieve these aims. In Section 4.3 evidence will be presented to demonstrate the extent to which these goals have been realized. For now, an indication of how language analysis relates to indexes is in order.

3. Indexing Theory

An indexing system may be characterized, after Landry (9), as in Figure 6.1. In this model, four major components are identified, the input-document space \mathcal{D}_i ; the analysis-document space, \mathcal{D}_a ; the index space, \mathcal{I} ; and the several indexes, I_1, I_2, \dots, I_n , which the system is capable of producing. The indexing system embodies a set of procedures or mappings from input documents to index(es). These procedures are part of the document space \mathcal{D}_a . They operate on input

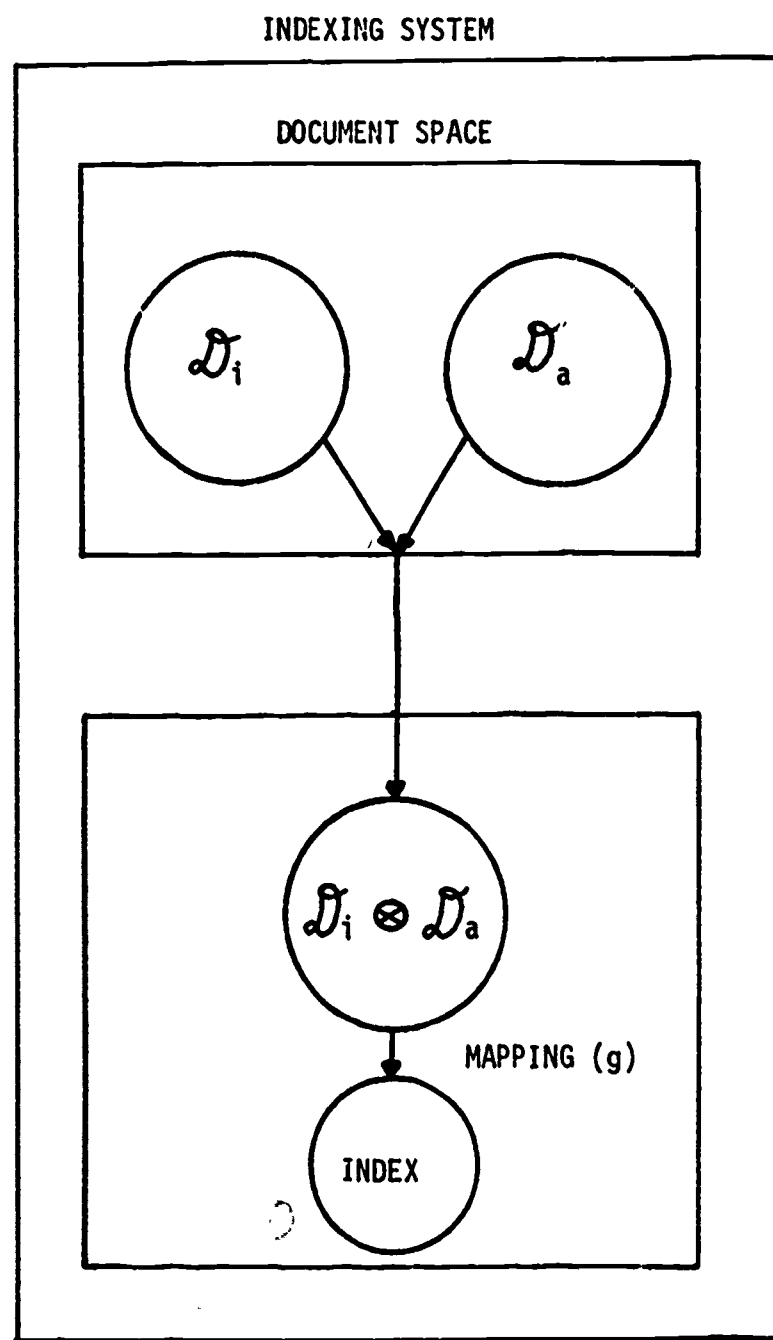


Figure 6.1 An indexing system as characterized by Landry (9).

documents first to create the document space \mathcal{D}_1 , then on the elements of \mathcal{D}_1 to produce the index space, and finally on the index space to produce an index. The index space is a space whose dimensionality is determined by the number of attributes the procedures in \mathcal{D}_1 ascribe to the input documents. In general, the higher the dimensionality of the index space the greater the variety and adequacy of the indexes the system can produce. Thus, the index space corresponds to a meta-representation of the input documents. The fidelity of this representation is a function of its dimensionality.

For the present research, the significance of these observations is that the analysis procedures described in Chapters III-V correspond to Landry's analysis documents and, since they identify a fairly large number of attributes of the input text, they may be supposed to produce an index space of high fidelity. While unequivocal evidence in support of this supposition cannot be given, its partial verification is discussed in the last part of this chapter. The index space may, I think, be viewed in terms of structural representations of language. The nature and derivation of such representations are discussed below.

4. Procedures for the Graphic Representation of Sentences

In the remainder of this chapter I describe a graphic representation of English sentences which I propose as an approximation to Landry's index space (10). The procedures outlined for generating this representation are, I believe, sufficient to permit their implementation with little difficulty. This chapter is not concerned, however, with the

exact methods for such implementation nor with the graph theory which might be pertinent. It is recognized that the mathematical theory involved in the storage, matching and construction of graphs is an important undertaking, but one whose magnitude is beyond the scope of this work.

4.1. Why a Graphic Representation?

It is well-known that English is not strictly linear even though it is produced in a linear fashion (11). This fact is easily illustrated by sentences containing embedded clauses, as

Sentences that contain embedded clauses have a non-linear structure.

For indexing, the import of this observation is that the specific relationships which are assigned to elements of a sentence are often not immediately obvious from the linear sequence that we call a sentence. But if an algorithmic way of explicating these relationships, that is, of producing a graphical representation of English text, could be devised it might be possible to derive indexes of various kinds from a single representation. Perhaps more important, the representation itself could serve as a kind of index which a person could access directly. Work most closely related to what I shall describe is that of Fugmann and his colleagues (12). This group has devised a machine-aided system for the production of graphic representation of patent literature which they believe will lead to considerably improved information retrieval systems. This work, as well as other related work is discussed in the next section.

4.2. Approaches to the Graphic Representation of Language

By way of preamble, it must be noted that the graphic representations which are of interest here are those involving directly the elements of a sentence. Hence, I exclude from consideration all those grammatical studies in which parse trees of various kinds are generated (or proposed).

Many researchers have recognized the value of graphic representation of sentences. Several approaches to the representation of language in graphic form are reviewed briefly below.

4.2.1. Graphic Structure Proposed by Salton

Salton has suggested graphs that represent sentences as an aid in each component of his information storage and retrieval system: analysis, identification, normalization and matching. Salton proposes representing sentences by using tree structures (a graph which has at most one branch entering each node and which contains no circuits (13)). The sentences are analyzed by using a dependency grammar in which the verb is the fulcrum, or central relator of the sentence (14). The graphs contain explicit syntactic relations such as noun-verb and noun-preposition. Semantic relations such as identity and location are also suggested as components of the graph. An example of such a graph is shown in Figure 6.2.

4.2.2. TOSAR Graphs

A somewhat different approach to the construction and use of graphs is that developed by Fugmann, Nickelsen, Nickelsen and Winter (12), involving the intellectual construction of graphs for use in the automated

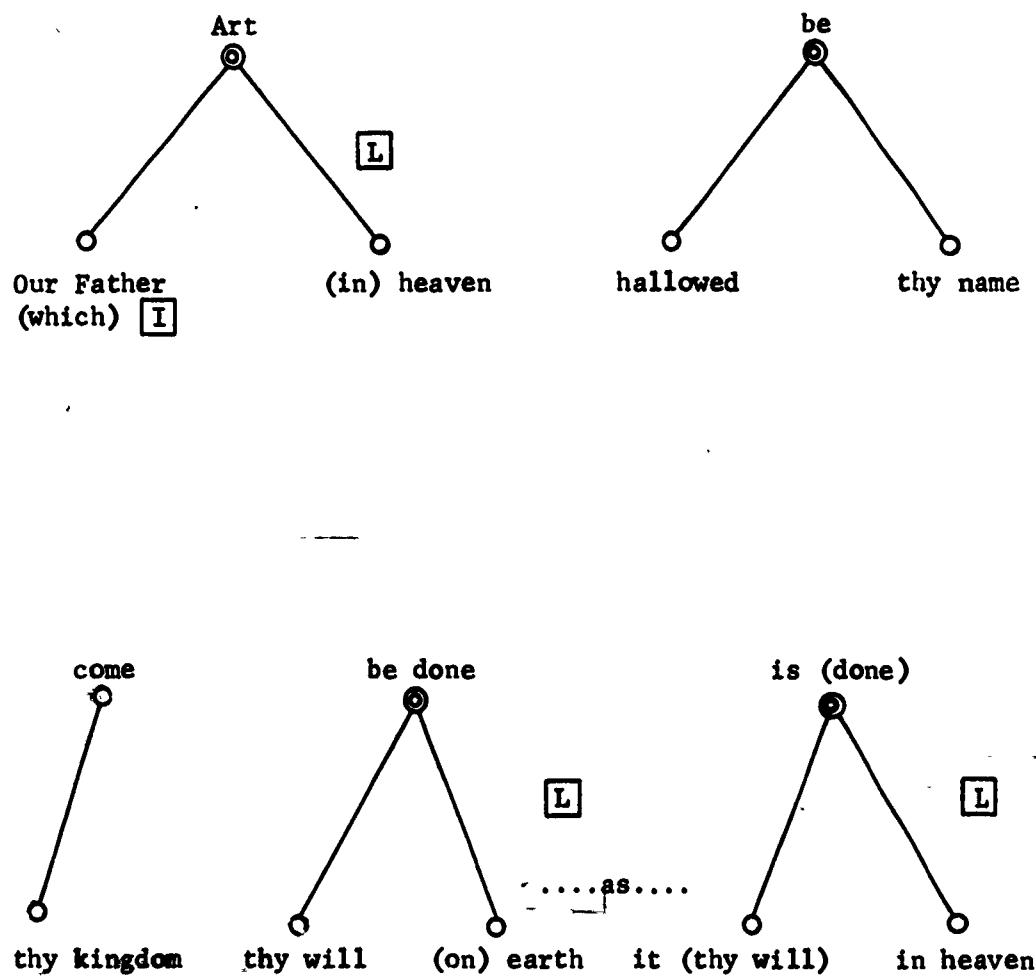
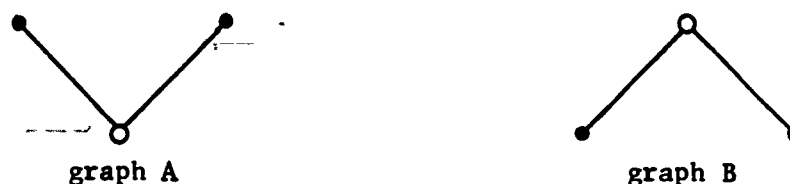


Figure 6.2

Graphic representation proposed by Salton.
 ([I] = Identity; [L] = Location) (15).

storage search and retrieval of concepts. The concepts used in the chemical literature lend themselves to graphic representation and hence provide a good starting point for investigation. In the graphic model TOSAR (Topological Method for the Representation of Synthetic Analytical and Relations of Concepts) developed by this group, relations between concepts are represented by lines which join the concepts. Nodes labelled with Roman letters represent substances and nodes labelled with a Greek letter represent processes. Each level of the graph represents a particular stage of a complex process. Thus the graph A characterizes the combination of two substances (by means of an unspecified process) while graph B characterizes the separation of a substance into two components (again without specifying the process). An example of a graph



produced via the TOSAR model is given in Figure 6.3.

4.2.3. Graphic Representation of Deep Structure

One purpose of the TOSAR model is the adequate representation of a concept which embodies several possible index entries. While chemical literature especially lends itself to this type of analysis, the developers of TOSAR believe that more investigation is needed to define all the concepts in a document in this way (16). An effort in this direction is the Conceptual Dependency Parser (CDP) developed by Schank and Tesler (17). The relevance of the CDP lies in its

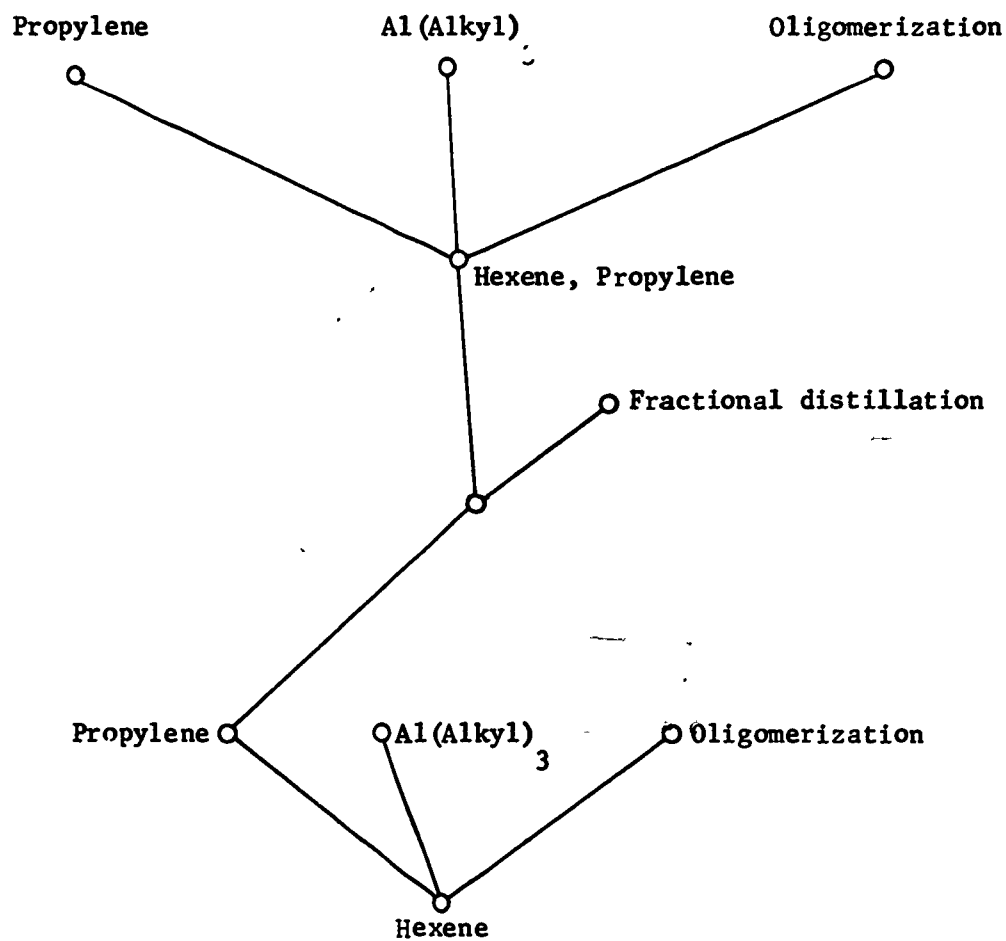


Figure 6.3

A graph produced via the TOSAR model of the sentence: Oligomerization of propylene with the aid of Al(Alkyl)₃ to obtain hexene, and separation and purification of the excess propylene by fractional distillation and recycling of the propylene (18).

similarity to case grammar and in its emphasis on the relational attributes of a sentence. According to Schank and Tesler, the CDP is an approach to the identification of the "underlying meaning" of a sentence. Noun phrases are assigned membership in a "governing" or an "assisting" class, and are assigned roles such as "actor" or "participant." Each word contained in the vocabulary of the parser is defined by a list of attributes similar to those suggested by Katz and Fodor (19). If two words occur which have incompatible or undefined attributes, the system must be updated to include the necessary relationship. An example obtained from the CDP is given in Figure 6.4 using the sentence "John's love is good." Each type of arrow (arc) depicts a different kind of relationship. Presumably, the sample sentence is broken down into two "concepts": "John loves" and "One (John's love) is good." The work of Schank and Tesler represents an attempt to acquire a more profound analysis of sentences than the current state of the art can provide.

4.2.4. Graphic Structures Proposed by Plath

Two other programs which must be mentioned have each been implemented for languages other than English. A program has been implemented which constructs diagrams of Russian sentences (20). The program is based on a projective grammar³³ and is used as a method of presenting the output of a predictive syntactic analyzer. The results have also been used to classify and analyze sentences according to their structural properties. An illustration of Plath's results is given in Figure 6.5.

33. A grammar similar to an immediate-constituent grammar.

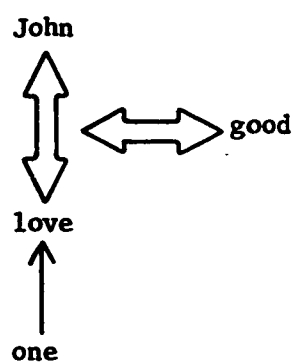


Figure 6.4 The "underlying meaning" of the sentence "John's love is good." as represented by the Conceptual Dependency Parser (17).

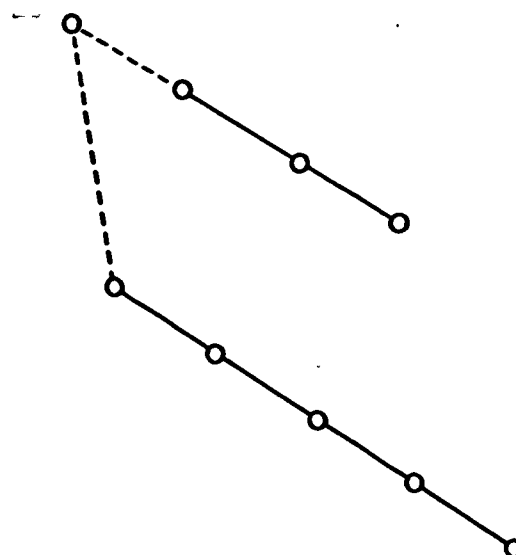
**AUTOMATICALLY GENERATED
SENTENCE DIAGRAM****IMPLIED NODAL
INTERCONNECTIONS****CLAUSE INDICATOR****NAPRJAZHENIJA****NA****KONDENSATORAX****OTSCHITYVAJUTSJA****NA****EHKRANE****OSTSILLOGRAFA****KATODNOGO**

Figure 6.5 Plath's Automatically Produced Sentence Diagrams (20).

4.2.5. Graphic Structures Proposed by Tesniere

The second program has been implemented for French sentences by Tesnieré (21). In this research the language is partitioned into four classes: verb, substantive, adjective and adverb. The verb is treated as the central relator of the sentence. Tesniere theorizes that the words of the sentence are more significant the lower they are on the graph. Thus, in the example sentence in Figure 6.6, "vert" and "libre" would be the most significant words.

4.3. A Proposal for the Production of Graphic Representations of English Sentences

4.3.1. Definition of Components

The production of a graphic representation of an English sentence hinges upon the relational elements of a sentence. A subgraph is produced for each string which contains a PRIMARY RELATION, and then the relationships between these subgraphs are established to produce a graph. Such a process could, in principle, be extended beyond sentence boundaries. The SECONDARY RELATIONS are represented by the arcs of the graph. All other elements are represented as nodes of the graph. COMPOSITE NAMES (i.e., NOMINAL PHRASES) are represented as single nodes for simplicity in presentation. The individual elements of a COMPOSITE NAME could be represented as distinct nodes of the graph, and the relationships could be indicated by the arcs which join these SIMPLE NAMES. The desirability of doing this depends upon the particular use intended for the graph.

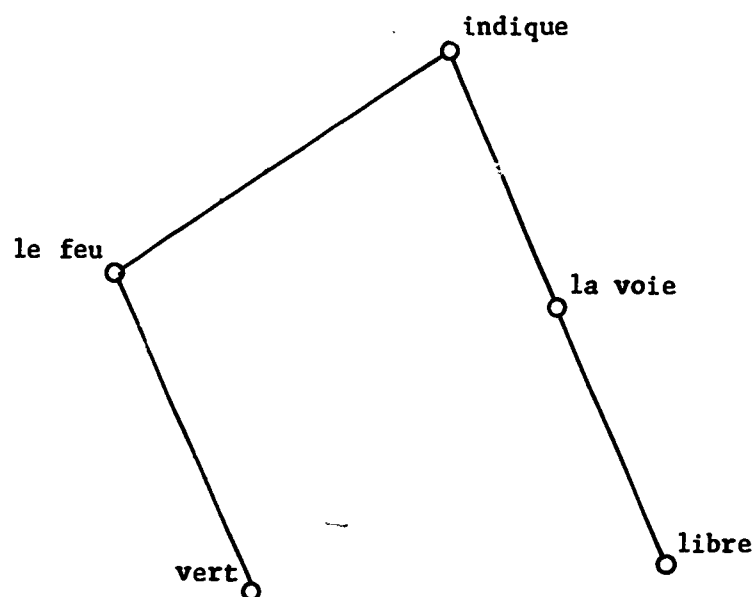


Figure 6.6 Tesnière analysis of the sentence "Le feu vert indique la voie libre." (The green light indicates clear passage.) (21)

Nodes of the graph will be distinguished as NAME nodes (N-nodes) and PRIMARY-RELATION nodes (PR-nodes). Again for simplicity of presentation, N-nodes will be labelled with capital Roman letters instead of the specific NAME, and PR-nodes will be specified by geometric shapes (see Table 6.1).

4.3.2. Representation of PRIMARY-RELATION Nodes

A PRIMARY RELATION will appear as a node (called a PR-node) in a graph, the form of which identifies the case-grammar class to which the PRIMARY RELATION belongs. Arcs extend from the PR-node in such a way that the direction of the link between the PRIMARY RELATION and its arguments (NAME) is indicated. If the PRIMARY RELATION is RECESSIVE, the symbol for the node indicates this fact. Specifically, PARTICIPIAL MAIN RECESSIVE PRIMARY RELATIONS are denoted by a darkened node and INFINITIVAL MAIN RPRs are denoted by a shaded node. Table 6.1 lists the node symbols which will be used for the various categories of both DOMINANT and RECESSIVE PRIMARY REALTIONS.

4.3.3. Representation of Case Types and COMPOSITE NAMES

The SECONDARY RELATIONS of a sentence are represented by the arcs of the graph. Since SECONDARY RELATIONS always have two arguments, the arc makes explicit the arguments of the RELATION. The direction of the arc will always be from the second argument to the first. The case role assignment to the second argument of a SECONDARY RELATION is retained to further specify this RELATION. The case role is indicated by a numeral over the arc. The numerical codes corresponding to the case types are given in Table 6.2. An example should clarify these

Table 6.1 Symbols Used to Represent Case Grammar Classes of PRIMARY RELATIONS.
















PRIMARY RELATION TYPE (CASE)	SYMBOLGY		
	DOMINANT	RECESSIVE	
		PARTICIPIAL	INFINITIVAL
Stative			
Agentive			
Experiencer			
Beneficiary			
Reflexive			

Table 6.2 The Case Roles Along with the Corresponding Numerical Codes Used in the Graphic Representation.

CASE ROLE	NUMERICAL CODE
Agent	1
Object	2
Experiencer	3
Beneficiary	4
Location	5
Time	6
Manner	7
Comitative	8
Cause	9
Purpose	10
Facet	0

issues. Suppose we have, in prefix notation (cf., Chapter II, Section 3.4.4),

$$h_2 \ AB$$

This would be represented in graphical form as

$$A \leftarrow B$$

Likewise, the sentence fragment

$$\underbrace{\text{The boy}}_A \quad h_2 \quad \underbrace{\text{in the red shirt}}_B \dots$$

would be represented graphically as

$$A \xleftarrow{0} B$$

The item "facet" listed along with the case types, is not a case type but indicates an unmarked adjectival relationship. Recall that a case is assigned to a SIMPLE or COMPOSITE NAME. A NAME which implicitly modifies a second NAME does not, therefore, receive a case grammar assignment. Also, it is important to note that the case role is that of the node from which an arc emanates.

4.3.4. Representation of Consecutive SECONDARY PHRASEs

A construction which presents difficulty in analysis is a COMPLEX NAME consisting of a succession of SECONDARY PHRASEs. The difficulty which this construction poses is that of correctly identifying the first argument of the pair required by the SECONDARY RELATION (recall that the second argument of the pair always immediately follows the

SECONDARY RELATION). Consider the sentence:

The exam was given to the girl in the library.

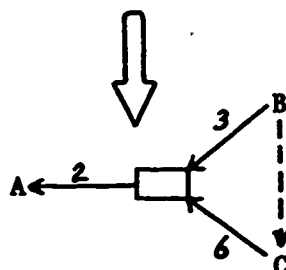
A b i₂ B i₂ C

For the first SECONDARY RELATION, i_2 the two required arguments are A and b_1 . But for i_2 it is not possible to determine unequivocally whether the arguments are b_1 and B or B and C. Traditional grammar would employ a proximity rule which states that a noun phrase modifies the noun or verb phrase to which it is nearest. While this rule might eliminate much confusion, it would place a severe restriction on the interpretation of the sentence. The solution I propose is the following. Since the PRIMARY RELATION is the central RELATION of a SENTENCE, all NAMES are related, at least indirectly, through the PRIMARY RELATION. All ADVERBIAL SECONDARY RELATIONS (see Section 3.4) will therefore take as their first arguments the PRIMARY RELATION of the CLAUSE in which the SECONDARY RELATIONS are found. But this somewhat arbitrary decision makes it desirable for the graphic structure to contain an indication of the linear order of the phrases in the sentence. Such an indication is provided by a dashed arrow drawn from the COMPOSITE NAME of a SECONDARY PHRASE to the COMPOSITE NAME of the SECONDARY PHRASE which immediately precedes it in the CLAUSE. For example,

The exam was given to the girl in the library.

The exam was given to the girl in the library

A
B
C



The dashed arrow is thus an ordering relation such that "B ----> C" implies B precedes C in the sentence from which the graph was derived.

4.3.5. Representation of PHRASES Related by COORDINATE CONJUNCTIVE SECONDARY RELATIONS

The particular elements of CCP found in a SENTENCE may be important to the interpretation of the SENTENCE. Thus, each member of CCP is represented in a graph as an arc labelled with a numeral (see Table 6.3.). In principle, there may be any number of PHRASES joined by an element of CCN. Each PHRASE which is part of such a series is linked to the next in succession by a dashed arc which indicates their sequential order. The following examples serve for illustration.

The meal consisted of cheese, bread and wine.

A
B
C
D

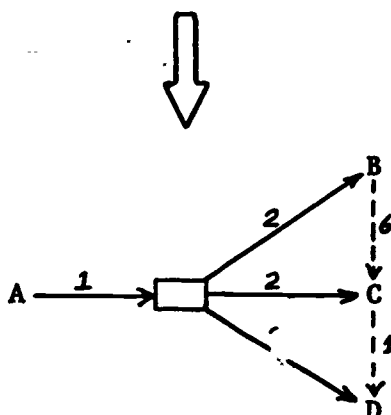


Table 6.3 COORDINATE CONJUNCTIVE SECONDARY RELATIONS and their Corresponding Numerical Codes.

Element	Numerical Code
and	1
or	2
but	3
nor	4
not	5
,	6

The dog ran across the field, over the bridge and down the lane.

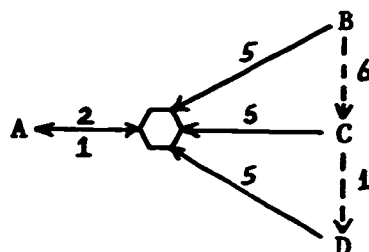
A



B

C

D



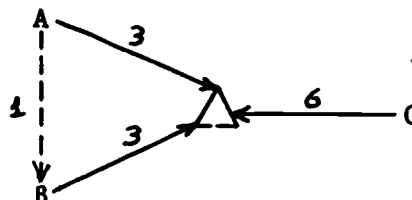
Students and faculty were hurt during the riots.

A

B



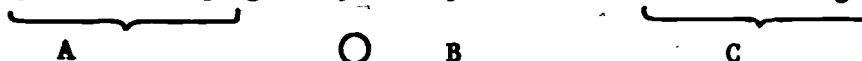
C



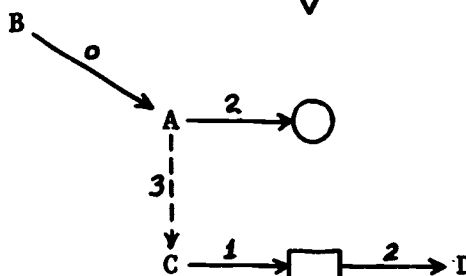
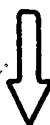
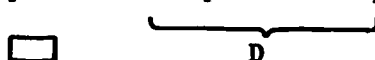
4.3.6. Representation of CLAUSES Joined by CEN

Clauses joined by element of CEN generally have a common element such as a PRIMARY RELATION or a NOMINAL PHRASE. If there is a common NOMINAL PHRASE, the right-most NAME of the first NOMINAL PHRASE of one CLAUSE is compared with the right-most NAME of the NOMINAL PHRASE of the second CLAUSE. If these NAMES are the same the CLAUSES are joined by an arrow connecting the NOMINAL PHRASES. Otherwise, an arc is drawn connecting the PRIMARY RELATION of each CLAUSE. Again, the direction of the arc indicates the order of the CLAUSES. The following examples illustrate this treatment.

High-level languages may be expensive but assembler language



requires more specialized personnel.



The men worked in the field and the women worked in the barn.

A

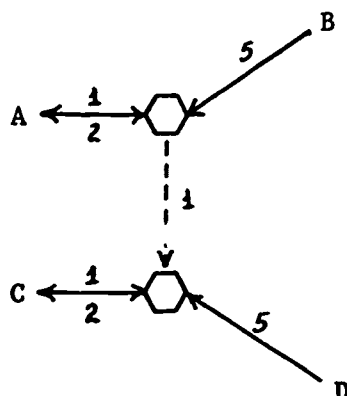


B

C



D



The boy jumped out the window and fell into the net.

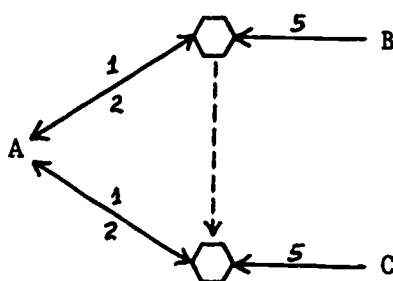
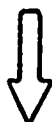
A



B



C

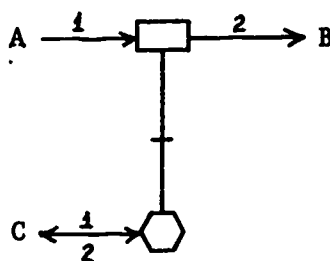
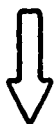


4.3.7. Representation of CLAUSES Joined by SCN

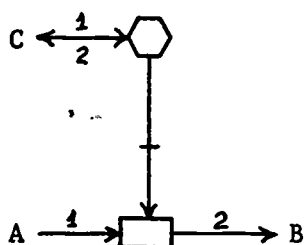
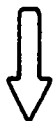
Two clauses which are joined by an element of SCN are connected by a marked arc ($\text{---}+\text{---}$). The arc is always drawn from the CLAUSE initiated by SCN to the second CLAUSE. If desired, the ordering of the CLAUSES is easily preserved. The following examples illustrate how CLAUSES joined by an element of SCN are depicted and how the ordering of CLAUSES can be preserved.

He finished the work after his friends left.

A  B  C



After his friends left he finished the work.

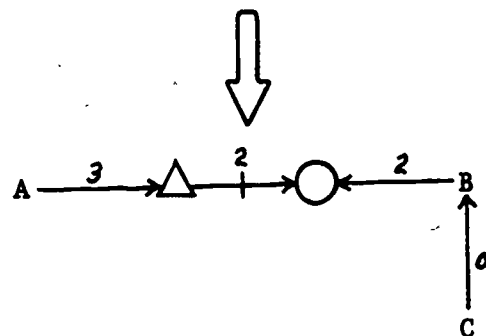


4.3.8. Representation of NOMINAL CLAUSES

A **CLAUSE** may serve in a **NOMINAL** position of a **PRIMARY CLAUSE**. When a **CLAUSE** serves as an argument of a **PRIMARY RELATION** in a second **CLAUSE**, a marked arc is drawn from the **PRIMARY RELATION** of the **PRIMARY CLAUSE** to the **PRIMARY RELATION** of the **NOMINAL CLAUSE**. The following examples are illustrative of this treatment. (Note that the **NOMINAL CLAUSE** may consist of just a **PRIMARY RELATION**).

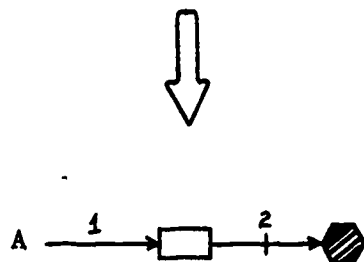
The boy knew that the answer was wrong.

A \triangle B \circ



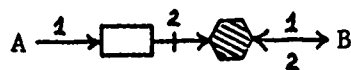
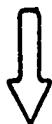
i would like to go.

A \square \bullet




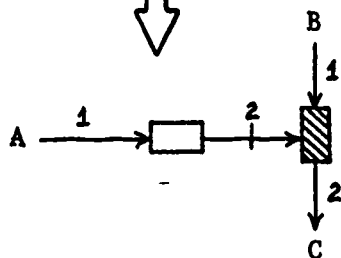
I would like him to go.

A ☐ B 



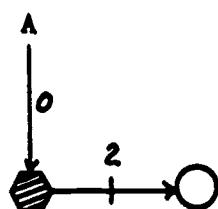
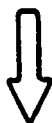
Mary wanted him to mail the letter.

A ☐ B  C



To err is human.

● ○ A



Effort on sentence classification has been devoted to finding

A

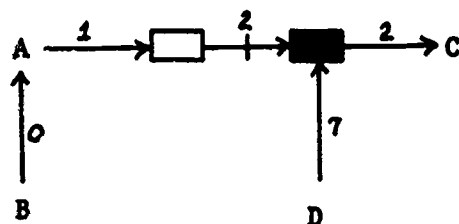
B



algorithms for implementation.

C

D

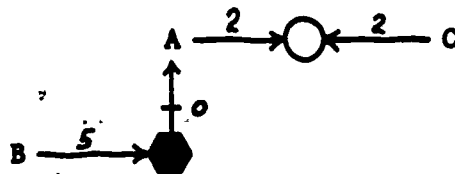


4.3.9. Representation of ADJECTIVAL CLAUSES

ADJECTIVAL CLAUSES are connected to a NOMINAL PHRASE by a marked arc. The arc is drawn from the PRIMARY RELATION of the ADJECTIVAL CLAUSE to the NOMINAL PHRASE.

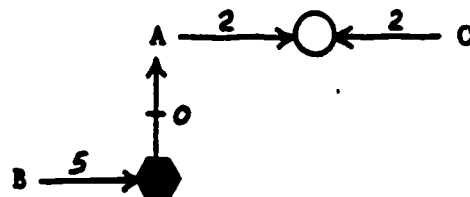
The girl who is standing in the hall is my daughter.

A
●
C
○
B



The girl standing in the hall is my daughter.

A ● B ○ C



Mary disliked John throwing rocks.

A □ B ■ C



4.3.10. Representation of ADVERBIAL CLAUSES

ADVERBIAL CLAUSES are depicted in a manner similar to that for NOMINAL and ADJECTIVAL CLAUSES. A marked arc is drawn from the PRIMARY RELATION of the ADVERBIAL CLAUSE to the PRIMARY RELATION which it modifies. Examples of adverbial clauses are:

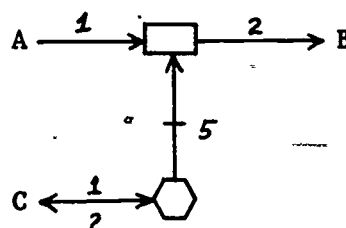
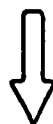
They rowed the boat to where the birds were circling.

A 

B

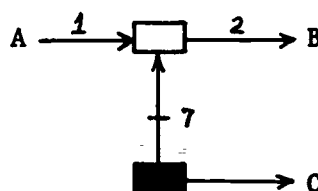
C





They tested the compounds by using the new equipment.

C  B  C



These examples serve to illustrate the way in which SENTENCES in English may be transformed into graphical representation. Such representations are susceptible of production by computer program. Although the programs for doing so have not been written, all of the necessary data to carry out the transformation from linear sequences to graphical representations are made available by the programs developed in this research.

5. Summary

In this chapter, it has been suggested that the proposed graphical representation of SENTENCES provide a strong basis for the derivation of indexes. The representations preserve and explicate the NAMES and RELATIONS in a SENTENCE and make it possible to extract NAMES at various levels of complexity for various purposes. For instance, keyword indexes, KWIC indexes, articulated indexes, and so on can still be

derived from the graphical representation, but more complex aggregates can also be retrieved. Such graphs make it possible for the first time to do what the chemist calls a "substructure search", that is, to isolate component parts of a structure related to the searcher in some meaningful (to him) way.

Much of the work in this chapter has gone on in parallel with that carried out by Strong (22), and the interested reader may wish to consult that work for additional information on graphical representations of text.

References

1. J. Gardin, "Semantic Analysis Procedures in the Sciences of Man," Social Science Information 8 (1) 17-45 (1969).
2. J. Howell, Discourse Concerning the Precedency of Kings, Printed for Rowland Reynolds, London, 1664, 219.
3. R. Fugmann, H. Nickelsen, I. Nickelsen and J. H. Winter, "TOSAR-- A Topological Method for Representation of Synthetic and Analytical Relations of Concepts Angewandte Chemie 09 (08) 589-595 (1970).
4. H. Skolnik, "The Multiterm Index: A New Concept in Information Storage and Retrieval," Journal of Chemical Documentation 10 (02) 81-84 (1970).
5. A. E. Petrarca and W. M. Lay, The Double-KWIC Coordinate Index: A New Approach for Preparation of High-Quality Indexes by Automatic Indexing Techniques", Journal of Chemical Documentation 9, 256-261 (1969).
6. J. E. Armitage and M. F. Lynch, "Articulation in the Generation of Subject Indexes by Computer," Journal of Chemical Documentation, 7 (3), 170-178 (1967).
7. M. R. Quillian, "Semantic Memory", in M. Minsky (ed.) Semantic Information Processing, The MIT Press, Cambridge, Mass., 1968.
8. R. Fugmann, et al., op. cit., 589-595.
9. B. C. Landry, "A Theory of Indexing: Indexing Theory as a Model for Information Storage and Retrieval," Computer and Information Science Research Center, CISRC-TR-71-13, The Ohio State University, 1971, p.53.
10. ibid., p.42.
11. D. G. Hayes, "Dependency Theory: A Formalism and Some Observations," Language 40 (4), 511-525 (1964).
12. Fugmann, et al., op. cit., 589-595.

13. G. Salton, "Manipulation of Trees in Information Retrieval" Communications of the ACM, 05 (02), 103-144 (1962).
14. Hayes, op. cit., 511-525.
15. Salton, op. cit., p. 138.
16. Fugmann, et al., op. cit., p. 592.
17. R. C. Shank and L. Tesler, "A Conceptual Dependency Parser for Natural Language," Statistic Methods in Linguistics 06, 31-51 (1970).
18. Fugmann, et al., op. cit., p. 594.
19. J. J. Katz and J. A. Fodor, "The Structure of a Semantic Theory," Language 39 (02), 1963.
20. W. Plath, "Automatic Sentence Diagramming" 1961 International Conference on Machine Translation of Languages and Applied Language Analysis, 1 (13), 176-208 (1961).
21. L. Tesnieré, Elements de Syntaxe Structural, Paris, Klenchsuck, quoted by R. Pages: in "Relational Aspects of Conceptualization in Message Analysis, from J. M. Perreault, "Proceedings of the International Symposium on Relational Factors in Classification" Information Storage and Retrieval, 3 (4), 177-410 (1967).

CHAPTER VII. CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

1. Conclusions

One of Murphy's laws has it that whatever one wishes to do he must always do something else first. So it has been with this research. The general aim of the work has been toward automated indexing. But before that could be done a great deal of work in computational linguistics had to be done first. Thus programs have been developed which provide for the analysis of English text on several levels. These programs are all efficient (at least in relation to any earlier programs written for the same purposes) and they are effective. Furthermore they have all been related within a theoretical framework of language and are therefore logically consistent. Perhaps the most important observation to be made is that all the procedures are more nearly free of semantic inferences than any other similar programs so far developed. And the fact that the case grammar program described in Chapter V is the first such program to be developed is also a significant achievement.

But all of the language analysis procedures are only a means to an end. That end has not been reached, but the structural representations proposed in Chapter VI are close to the desired goal, namely the automatic generation of a relationally rich base from which better indexes might be derived. Further comment on that point is, unfortunately, idle speculation now.

Some significant progress has been made in language analysis by computer, and it may be hoped that these results will lead to new, improved indexing procedures in the near future.

2. Directions for Future Research

Several lines of investigation suggest themselves. Refinement of the language analysis procedures is one of these, one which is already underway (Chapter IV). Refinements may take two directions. One is to improve the efficiency of the programs by redesigning them or by recoding them in assembler language, for instance. The second is to improve the accuracy of the output of this procedure by adding or modifying rules and by further refining the dictionaries.

A second line of investigation concerns the completion of the definition and implementation of the graphical-representation procedures. This area too is under active study, but much needs to be done.

A third area of study is that dealing with more of the interpretational aspects of language. Thus it may be important not only to know how words relate to one another within a body of text, but also how they are related to other bodies of text or to "real-world" referents. For information retrieval systems this is an important question.

There are many other possibilities. The last to be mentioned is automatic abstracting. It may be expected that improved abstracting procedures could be developed using data derived by the language

analysis procedures. The precise way in which linguistic data would be used in automatic abstracting remains to be determined.

APPENDIX A

APPENDIX A. A LISTING OF DICTIONARY ELEMENTS AND THEIR FREQUENCY

The elements contained within the dictionary were chosen from the Brown University statistical analysis of a corpus of a million words (1). Elements were chosen on the basis of their frequency of occurrence. The extended dictionary consists of the entire listing presented in this Appendix. The basic dictionary consists of all the elements except for VRBs and ADVs.

ADV

<u>Element</u>	<u>Frequency</u>
only	1747
even	1171
also	1069
just	872
too	832
still	782
here	750
again	578
once	499
always	458
away	456
less	438
almost	432
later	397
often	368
early	366
perhaps	307
today	284
really	275
already	273
together	267
probably	261
ago	246
sometimes	221
further	218
usually	206
soon	199
near	198
alone	195
finally	191
farther	183
simply	170
actually	166
especially	160
certainly	143
ready	143
directly	141
particularly	146
likely	151
suddenly	153
nearly	141
merely	135
generally	132
clearly	128

<u>Element</u>	<u>Frequency</u>
somewhat	127
apparently	125
immediately	123
recently	123
daily	122
forward	115
slowly	115
obviously	114
completely	110
ahead	109
easily	106
hardly	106
exactly	103

AUX

<u>Element</u>	<u>Frequency</u>
is	10099
was	9816
be	6377
had	5133
are	4393
have	3941
were	3284
been	2472
has	2439
being	712
it's	302
having	279
I'm	268
am	228
that's*	186
can't	796
wasn't	154
there's	109
you're	151
hadn't	99
he'd	98
isn't	77
she'd	67
one's	65
they're	65
we're	58

MOD

<u>Element</u>	<u>Frequency</u>
would	2714
will	2244
can	1772
could	1599
may	1400
must	1013
should	888
might	672
shall	267
cannot	258
I'll	181
couldn't	175
wouldn't	129
won't	105
we'll	64
he'll	
she'll	

AJN

<u>Element</u>	<u>Frequency</u>
do	1363
did	1044
don't	489
does	485
didn't	401
doesn't	87
got	482
get	750
let	384

CCN

<u>Element</u>	<u>Frequency</u>
and	28852
but	4381
or	4207
nor	195
not	

SCN

<u>Element</u>	<u>Frequency</u>
if	2199
than	1789
then	1377
like	1290
since	628
however	552
though	442
yet	419
although	319
thus	312
whether	286
therefore	205
unless	101

DTR

<u>Element</u>	<u>Frequency</u>
the	69971
a	23237
his	6997
an	3747
their	2670
its	1858
my	1319
our	1252
your	923
1	496
every	491
2	450
3	282
4	196
third	193
10	143
5	134
6	113
15	109
30	106
8	104
12	98

DTR/PRN

<u>Element</u>	<u>Frequency</u>
this	5146
one	3292
her	3037
all	3001
more	2216
other	1702
some	1617
these	1573
two	1412
first	1360
any	1345
most	1160
many	1030
much	937
each	877
those	850
own*	772
both	730
same	686
another	683
three	610
few	601
enough	436
several	377
second	373
four	359
kind	313
five	286
six	220
million	204
hundred	171
ten	165
neither	141
ones	116
eight	104
whole	309
either	284

INT

<u>Element</u>	<u>Frequency</u>
very	796
rather	373
quite	281

PRP

<u>Element</u>	<u>Frequency</u>
of	36411
to	26149
in	21341
for	9489
with	7289
on	6742
at	5378
by	5305
from	4369
out	2096
up	1895
over	1236
after	1070
before	1016
through	969
down	895
between	730
under	707
off	639
during	585
without	583
around	561
upon	495
until	461
toward	386
among	370
within	359
along	355
above	296
across	282
outside	210
except	181
beyond	175
inside	174
instead	173
throughout	141
dispite	104
about	1815
into	1791
below	145
according	139
behind	258

PRN₁

<u>Element</u>	<u>Frequency</u>
he	9543
it	8756
I	5173
they	3618
you	3286
she	2859
we	2653
him	2619
them	1789
me	1181
us	612
things	368
least	343
thing	333
others	323
anyone	140
none	108

PRN₂

<u>Element</u>	<u>Frequency</u>
something	450
nothing	412
anything	280
everything	185

PRN₃

<u>Element</u>	<u>Frequency</u>
himself	603
itself	304
themselves	270
myself	129
herself	125

REL

<u>Element</u>	<u>Frequency</u>
which	3562
who	2252
what	1908
whose	252
whom	146
when	2231
where	938
how	834
while	680
why	404
whatever	112

VRB

<u>Element</u>	<u>Frequency</u>
said	1961
made	1125
make	794
see	772
know	683
come	630
go	625
came	622
take	611
found	536
went	507
say	504
put	437
think	433
took	426
set	414
told	413
find	399
going	399
look	399
knew	395
give	391
given	377
become	361
saw	352
want	329
done	320
began	312

<u>Element</u>	<u>Frequency</u>	<u>Element</u>	<u>Frequency</u>
gave	285	continue	107
taken	281	determine	107
seen	279	serve	107
tell	268	stress	107
held	264	applied	106
keep	264	closed	106
seems	258	reach	106
brought	253	write	105
heard	247	married	105
became	246	remained	105
known	245	covered	104
seem	229	played	104
provide	216	spent	104
believe	200	built	103
says	200	becomes	102
gone	195	related	102
kept	186	rise	102
wrote	181	meant	100
lead	173		
makes	172		
tried	170		
shown	166		
bring	158		
written	154		
hear	153		
sat	150		
meet	148		
paid	145		
sent	145		
showed	141		
remember	138		
comes	137		
understand	132		
ran	134		
led	132		
met	132		
ask	128		
consider	126		
appear	118		
born	113		
include	113		
gives	112		
speak	110		
expect	108		

APPENDIX B

References

1. H. Kucera and W. R. Francis, Computational Analysis of Present-Day American English, Brown University, Providence, R. I., 1967.

APPENDIX B. A LISTING OF THE SOURCE DOCUMENTS FOR THE TESTING OF
COMPUTER PROGRAMS.

The documents are:

"The Need for a More Precise Definition of 'Algorithm'"

The Old Man and the Sea

"The Clavichord and How to Play It"

THE NEED FOR A MORE PRECISE DEFINITION OF "ALGORITHM". #3. A. TRAKHTENBROT, ALGORITHMS AND AUTOMATIC COMPUTING MACHINES, 52-57 (1962). OUR PREVIOUS DISCUSSION SHOWS THE STRONG CONNECTION WHICH EXISTS BETWEEN ALGORITHMS AND AUTOMATIC COMPUTING MACHINES. OBVIOUSLY, ANY PROCESS WHICH CAN BE PERFORMED BY A MACHINE CAN BE WRITTEN AS AN ALGORITHM. CONVERSELY, ALL ALGORITHMS WHICH HAVE SO FAR BEEN CONSTRUCTED, AS WELL AS THOSE WHICH MAY BE EXPECTED IN THE PRESENT STATE OF SCIENCE, CAN IN PRINCIPLE BE PERFORMED BY MACHINE. THE LAST STATEMENT REQUIRES SOME CLARIFICATION. AS WE HAVE ALREADY SEEN, THE ACTUAL APPLICATION OF AN ALGORITHM MAY TURN OUT TO BE VERY LENGTHY, AND THE JOB OF RECORDING ALL OF THE INFORMATION INVOLVED MAY BE ENORMOUS. ON THE OTHER HAND, THE MEMORY UNITS OF MACHINES HAVE A LIMITED CAPACITY (SINCE THE NUMBER OF MEMORY CELLS IS FINITE AND THE CAPACITY OF EACH CELL IS LIMITED). THEREFORE, IT MAY TURN OUT TO BE IMPOSSIBLE TO EXECUTE AN ALGORITHM UNDER EXISTING CONDITIONS. THIS CAN BE ILLUSTRATED BY THE EUCLIDEAN ALGORITHM. THE VERY SIMPLE PROBLEM OF FINDING THE GREATEST COMMON DIVISOR OF TWO NUMBERS CANNOT BE SOLVED BY HAND IF IT REQUIRES MORE PAPER AND INK THAN IS AVAILABLE. SIMILARLY, A PROBLEM WILL NOT BE SOLVABLE BY MACHINE IF IT REQUIRES MORE MEMORY SPACE THAN THERE IS IN THE MACHINE. IN SUCH CASES WE SAY THAT AN ALGORITHM IS POTENTIALLY REALIZABLE IF IT LEADS TO THE REQUIRED RESULT IN A FINITE NUMBER OF STEPS (EVEN THOUGH THIS NUMBER MAY BE VERY LARGE). IN OTHER WORDS, IT WOULD BE POSSIBLE TO USE THE ALGORITHM IN A MACHINE WHICH HAD AN UNLIMITED MEMORY CAPACITY. THE CONNECTION BETWEEN THE IDEA OF AN ALGORITHM AND THE IDEA OF AN AUTOMATIC MACHINE WITH A MEMORY OF INFINITE CAPACITY LEADS TO A CLEAR UNDERSTANDING OF THE NATURE OF EACH. HOWEVER, FOR ALL OF OUR EMPHASIS ON THEIR CONNECTION, WE STILL HAVE NOT DEFINED EITHER OF THESE IDEAS PRECISELY AN EXACT MATHEMATICAL DEFINITION OF THE NOTION OF ALGORITHMS (AND, AT THE SAME TIME, OF AUTOMATIC COMPUTING MACHINES) WAS NOT PRODUCED UNTIL THE 1930'S. WHY, THROUGH THE COURSE OF MANY CENTURIES, HAVE MATHEMATICIANS TOLERATED WITHOUT ANY PARTICULAR QUALMS AN UNCLEAR NOTION OF ALGORITHMS? WHY IS IT THAT ONLY RECENTLY HAS AN ACUTE NEED FOR A DEFINITION SUFFICIENTLY EXACT FOR MATHEMATICAL DISCUSSION ARISEN? EARLIER, THE TERM "ALGORITHM" OCCURRED IN MATHEMATICS ONLY IN CONNECTION WITH CONCRETE ALGORITHMS, WHERE AN ASSERTION OF THE EXISTENCE OF AN ALGORITHM WAS ALWAYS ACCOMPANIED BY A DESCRIPTION OF SUCH AN ALGORITHM. UNDER THESE CONDITIONS IT WAS NECESSARY TO SHOW ONLY THAT THE SYSTEM OF FORMAL INSTRUCTION WHEN APPLIED TO ANY DATA IN FACT LED AUTOMATICALLY TO THE DESIRED RESULT. THUS, THE NEED FOR A PRECISE DEFINITION OF THE NOTION OF ALGORITHM NEVER AROSE, ALTHOUGH EVERY MATHEMATICIAN HAD A WORKING IDEA OF WHAT THE TERM MEANT. HOWEVER, IN THE COURSE OF MATHEMATICAL PROGRESS, FACTS BEGAN TO ACCUMULATE WHICH RADICALLY CHANGED THE SITUATION. THE MOTIVATING FORCE WAS THE NATURAL DESIRE OF MATHEMATICIANS TO CONSTRUCT INCREASINGLY POWERFUL ALGORITHMS FOR SOLVING INCREASINGLY GENERAL TYPES OF PROBLEMS. RECALL THE ALGORITHM FOR FINDING SQUARE ROOTS. WE MIGHT WISH TO GENERALIZE THIS PROBLEM: TO CONSTRUCT AN ALGORITHM FOR FINDING THE ROOT OF ANY DEGREE OF ANY GIVEN NUMBER. IT IS NATURAL TO EXPECT THAT SUCH AN ALGORITHM WILL BE MORE DIFFICULT TO CONSTRUCT, BUT THE PROSPECT OF HAVING IT IS ATTRACTIVE. WE MAY GO EVEN FURTHER. FINDING THE N TH ROOT OF A NUMBER A MEANS SOLVING THE EQUATION $x^{*N} - A = 0$ (FINDING THE ROOTS OF THE EQUATION). WE CAN FORMULATE THE STILL MORE GENERAL PROBLEM: CONSTRUCT AN ALGORITHM FOR FINDING ALL ROOTS OF AN N TH DEGREE EQUATION OF THE FORM $A(N)x^{*N} + A(N-1)x^{*(N-1)} + \dots + A(1)x + A(0) = 0$, WHERE N IS AN ARBITRARY POSITIVE INTEGER. THE CONSTRUCTION OF SUCH AN ALGORITHM IS STILL MORE DIFFICULT. IN FACT, THE BASIC CONTENT OF THE THEORY OF EQUATION AMOUNTS TO THE CONSTRUCTION OF JUST THIS ALGORITHM; IT IS OF THE GREATEST IMPORTANCE. THE EXAMPLES GIVEN SHOW THE NATURAL STRIVING OF MATHEMATICIANS TO FIND INCREASINGLY POWERFUL ALGORITHMS TO SOLVE INCREASINGLY GENERAL TYPES OF PROBLEMS. OF COURSE, THE EXAMPLE OF SOLVING ALL EQUATIONS OF THE FORM ABOVE DOES NOT REPRESENT THE LIMIT TO WHICH ONE MIGHT GO. IF WE WANT TO PUSH THIS DESIRE FOR INCREASINGLY GENERAL ALGORITHMS TO THE EXTREME WE MUST INEVITABLY CONSIDER THIS PROBLEM: CONSTRUCT AN ALGORITHM FOR SOLVING ANY MATHEMATICAL PROBLEM. THIS IS A PROBLEM SO GENERAL THAT IT MIGHT BE CONSIDERED AN INSOLUBLE CHALLENGE TO MATHEMATICS AS A WHOLE. BESIDES THIS, IT CAN BE CRITICIZED ON THE GROUNDS THAT IT IS NOT CLEAR WHAT IS MEANT BY "ANY MATHEMATICAL PROBLEM." AT THE SAME TIME, THE GREAT ALLURE OF SOLVING SUCH A PROBLEM CANNOT BE DOUBTED. THIS PROBLEM HAS ITS OWN HISTORY. THE GREAT GERMAN MATHEMATICIAN AND PHILOSOPHER LEIBNIZ (1646-1716) DREAMED OF AN ALL-INCLUSIVE METHOD FOR SOLVING ANY PROBLEM. ALTHOUGH HE WAS UNABLE TO FIND IT, LEIBNIZ STILL THOUGHT THAT THE TIME WOULD

WULD COME WHEN IT WOULD BE DISCOVERED, AND THAT ANY ARGUMENT AMONG MATHEMATICIANS COULD THEN AUTOMATICALLY BE SETTLED WITH PENCIL AND PAPER. LATER, THE PROBLEM RECEIVED SOME RELIEFMENT IN THE FORM OF ONE OF THE MOST FAMOUS PROBLEMS OF MATHEMATICAL LOGIC, THE DECIDABILITY PROBLEM. SINCE WE DO NOT HAVE THE ROOM FOR A COMPLETE TREATMENT OF THE PROBLEM, WE SHALL MERELY SKETCH ITS GENERAL OUTLINES. AS IS WELL KNOWN, THE AXIOMATIC METHOD IN MATHEMATICS CONSISTS IN DERIVING ALL THE THEOREMS IN A GIVEN THEORY BY FORMAL LOGICAL STEPS FROM CERTAIN AXIOMS WHICH ARE ACCEPTED WITHOUT PROOF. THE FIRST OF ALL AXIOMATIC THEORIES WAS GEOMETRY, BUT IN MODERN MATHEMATICS ALMOST ALL THEORIES ARE CONSTRUCTED AXIOMATICALLY. MATHEMATICAL LOGIC EMPLOYS A SPECIAL "LANGUAGE OF FORMULAS" THAT ENABLES US TO WRITE ANY PROPOSITION OF A MATHEMATICAL THEORY AS A UNIQUELY DETERMINED FORMULA. IN THE TERMINOLOGY WHICH WE USED EARLIER FOR AN ASSOCIATIVE CALCULUS, WE MAY SAY THAT SUCH A FORMULA IS A WORD IN A SPECIAL ALPHABET CONTAINING SYMBOLS TO DENOTE LOGICAL OPERATIONS SUCH AS NEGATION, CONJUNCTION, AND IMPLICATION, AS WELL AS THE USUAL MATHEMATICAL SYMBOLS, SUCH AS PARENTHESES, AND LETTERS TO DENOTE FUNCTIONS AND NO VARIABLES. HOWEVER, THE CHIEF SIMILARITY TO AN ASSOCIATIVE CALCULUS CONSISTS IN THE POSSIBILITY OF WRITING THE LOGICAL DERIVATION OF A STATEMENT S FROM A PREMISE R IN THE FORM OF FORMAL TRANSFORMATION OF WORDS, VERY SIMILAR TO THE ADMISSIBLE SUBSTITUTIONS IN AN ASSOCIATIVE CALCULUS. THIS ALLOWS US TO SPEAK OF A LOGICAL CALCULUS, WITH A SYSTEM OF ADMISSIBLE TRANSFORMATIONS REPRESENTING ELEMENTARY ACTS OF LOGICAL DEDUCTION, FROM WHICH ANY LOGICAL INFERENCE, OF ARBITRARY COMPLEXITY, MAY BE BUILT. AN EXAMPLE OF SUCH AN ADMISSIBLE TRANSFORMATION IS THE ELIMINATION OF TWO CONSECUTIVE NEGATIONS IN A FORMULA; THUS, "NOT UNPROVED" MAY BE TRANSFORMED INTO "PROVED." THE QUESTION OF THE LOGICAL DEDUCIBILITY OF THE PROPOSITION S FROM THE PREMISE R IN A LOGICAL CALCULUS BECOMES THE QUESTION OF THE EXISTENCE OF A DEDUCTIVE CHAIN LEADING FROM THE WORD REPRESENTING R TO THE WORD REPRESENTING S . THE DEDUCIBILITY PROBLEM MAY NOW BE FORMULATED AS FOLLOWS: FOR ANY TWO WORDS (FORMULAS) R AND S IN A LOGICAL CALCULUS, DETERMINE WHETHER OR NOT THERE EXISTS A DEDUCTIVE CHAIN FROM R TO S . THE SOLUTION IS SUPPOSED TO BE AN ALGORITHM FOR SOLVING ANY PROBLEM OF THIS TYPE (ANY R AND S). SUCH AN ALGORITHM WOULD GIVE A GENERAL METHOD FOR SOLVING PROBLEMS IN ALL MATHEMATICAL THEORIES WHICH ARE CONSTRUCTED AXIOMATICALLY (OR RATHER, IN ALL FINITELY AXIOMATIZABLE THEORIES). THE VALIDITY OF ANY STATEMENT S IN SUCH A THEORY MERELY MEANS THAT IT CAN BE DEDUCED FROM THE SYSTEM OF AXIOMS, OR WHAT IS THE SAME THING, THAT IT CAN BE DEDUCED FROM THE STATEMENT R WHICH ASSERTS THAT ALL THE AXIOMS HOLD. THEN THE APPLICATION OF THE ALGORITHM WOULD DETERMINE WHETHER OR NOT THE PROPOSITION S WERE VALID. MOREOVER, IF THE PROPOSITION S WERE VALID, THEN WE COULD FIND A CORRESPONDING DEDUCTIVE CHAIN OF INFERENCE WHICH WOULD PROVE THE PROPOSITION. THE PROPOSED ALGORITHMS WOULD IN FACT BE A SINGLE EFFECTIVE METHOD FOR SOLVING ALMOST ALL OF THE MATHEMATICAL PROBLEMS WHICH HAVE BEEN FORMULATED AND REMAIN UNSOLVED TO THIS DAY. THAT IS WHY CONSTRUCTING SUCH AN "ALL-INCLUSIVE ALGORITHM" AND AN "OMNIPOTENT MACHINE" TO MATCH IT IS SO APPEALING A PROSPECT AND AT THE SAME TIME SO DIFFICULT. FINDING SUCH AN ALGORITHM HAVE REMAINED INSURMOUNTABLE. FURTHERMORE, SIMILAR DIFFICULTIES WERE SOON ENCOUNTERED IN TRYING TO FIND ALGORITHMS FOR CERTAIN PROBLEMS OF A FAR LESS GENERAL NATURE. AMONG THESE WERE HILBERT'S PROBLEM ON DIOPHANTINE EQUATIONS, AS WELL AS OTHERS WHICH WILL BE DISCUSSED BELOW. AS A RESULT OF MANY FRUITLESS ATTEMPTS TO CONSTRUCT SUCH ALGORITHMS, IT BECAME CLEAR THAT THE DIFFICULTIES INVOLVED ARE BASIC, AND IT CAME TO BE SUSPECTED THAT IT IS NOT POSSIBLE TO CONSTRUCT AN ALGORITHM FOR EVERY CLASS OF PROBLEMS. THE ASSERTION THAT A CERTAIN CLASS OF PROBLEMS CANNOT BE SOLVED ALGORITHMICALLY IS NOT SIMPLY A STATEMENT THAT NO ALGORITHM HAS YET BEEN DISCOVERED. IT IS THE STATEMENT THAT SUCH AN ALGORITHM IN FACT CAN NEVER BE DISCOVERED, IN OTHER WORDS, THAT NO SUCH AN ALGORITHM CAN EXIST. THIS ASSERTION MUST BE BASED ON SOME SORT OF MATHEMATICAL PROOF; HOWEVER, SUCH A PROOF MAKES NO SENSE UNTIL WE HAVE A PRECISE DEFINITION OF "ALGORITHM", SINCE UNTIL THEN IT IS NOT CLEAR WHAT IT IS WE ARE TRYING TO PROVE IMPOSSIBLE. IT IS USEFUL TO REMEMBER AT THIS POINT THAT IN THE HISTORY OF MATHEMATICS THERE HAVE BEEN OTHER PROBLEMS FOR WHICH SOLUTIONS HAD BEEN SOUGHT IN VAIN FOR A LONG TIME, AND FOR WHICH IT WAS ONLY LATER PROVED THAT SOLUTIONS COULD NOT BE OBTAINED. EXAMPLES ARE THE PROBLEM OF TRISECTING THE ANGLE AND THE PROBLEM OF SOLVING THE GENERAL FIFTH DEGREE EQUATION BY RADICALS. A METHOD OF TRISECTING AN ANGLE USING COMPASS AND STRAIGHTEDGE IS KNOWN TO EVERY SCHOOLBOY. THE ANCIENT GREEKS TRIED TO SOLVE THE PROBLEM OF TRISECTING AN ANGLE

OF USING COMPASS AND STRAIGHTEDGE. IT WAS LATER PROVED THAT TRISECTION OF AN ARBITRARY ANGLE BY SUCH MEANS IS IMPOSSIBLE. IT IS ALSO WELL KNOWN THAT THE SOLUTION OF A QUADRATIC EQUATION CAN BE EXPRESSED IN TERMS OF THE COEFFICIENTS BY MEANS OF A FORMULA WHICH EMPLOYS THE SIGNS FOR THE ARITHMETIC OPERATIONS AND THE RADICAL SIGN. THERE ARE ALSO FORMULAS IN RADICALS, WHICH ARE EXTREMELY COMPLICATED.

FOR THIRD AND FOURTH DEGREE EQUATIONS, A SEARCH FOR SIMILAR SOLUTIONS BY RADICALS FOR EQUATIONS OF DEGREE HIGHER THAN FOUR WAS CARRIED ON UNSUCCESSFULLY UNTIL THE BEGINNING OF THE NINETEENTH CENTURY, WHEN THE FOLLOWING REMARKABLE RESULT WAS FINALLY ESTABLISHED. FOR ANY n GREATER THAN OR EQUAL TO FIVE, IT IS IMPOSSIBLE TO EXPRESS THE ROOTS OF THE GENERAL n TH DEGREE EQUATION IN TERMS OF ITS COEFFICIENTS BY MEANS OF THE ARITHMETIC OPERATIONS AND THE OPERATION OF EXTRACTING n ROOTS. IN BOTH THESE CASES THE PROOF OF IMPOSSIBILITY TURNED OUT TO BE FEASIBLE ONLY AFTER THERE WERE PRECISE DEFINITIONS TO ANSWER THE QUESTIONS "WHAT IS MEANT

BY A COMPASS AND STRAIGHTEDGE CONSTRUCTION?" AND "WHAT IS MEANT BY SOLVING AN EQUATION IN RADICALS?" NOTE THAT THESE TWO DEFINITIONS GAVE A MORE PRECISE MEANING TO CERTAIN SPECIAL ALGORITHMS, NAMELY, THE ALGORITHM FOR SOLVING AN EQUATION IN RADICALS (NOT FOR THE SOLUTION OF EQUATIONS IN GENERAL) AND THE ALGORITHM FOR TRISECTING AN ANGLE WITH COMPASS AND STRAIGHTEDGE (NOT FOR TRISECTING BY ARBITRARY DEVICES). UNTIL RECENTLY, THERE WAS NO PRECISE DEFINITION OF THE CONCEPT "ALGORITHM" AND THEREFORE THE CONSTRUCTION OF SUCH A DEFINITION CAME TO BE ONE OF THE MAJOR PROBLEMS OF MODERN MATHEMATICS. IT IS VERY IMPORTANT TO POINT OUT THAT THE FORMULATION OF A DEFINITION OF "ALGORITHM" (OR OF ANY OTHER MATHEMATICAL DEFINITION) MUST BE CONSIDERED NOT MERELY AN ARBITRARY AGREEMENT AMONG MATHEMATICIANS AS TO WHAT THE MEANING OF THE WORD "ALGORITHM" SHOULD BE. THE DEFINITION HAS

TO REFLECT ACCURATELY THE SUBSTANCE OF THOSE IDEAS WHICH ARE ACTUALLY HELD, HOWEVER VAGUELY, AND WHICH HAVE ALREADY BEEN ILLUSTRATED BY MANY EXAMPLES. WITH THIS AIM, A SERIES OF INVESTIGATIONS WAS UNDERTAKEN BEGINNING IN THE 1930'S FOR CHARACTERIZING ALL THE METHODS WHICH WERE ACTUALLY USED IN CONSTRUCTING ALGORITHMS.

THE PROBLEM WAS TO FORMULATE A DEFINITION OF THE CONCEPT OF ALGORITHM WHICH WOULD BE COMPLETE NOT ONLY IN FORM, BUT, MORE IMPORTANT, IN SUBSTANCE. VARIOUS WORKERS PROCEEDED FROM DIFFERENT LOGICAL STARTING POINTS, AND BECAUSE OF THIS, SEVERAL DEFINITIONS WERE PROPOSED. HOWEVER, IT TURNED OUT THAT ALL OF THESE WERE EQUIVALENT, AND THEY DEFINED THE SAME CONCEPT; THIS WAS THE MODERN DEFINITION OF ALGORITHM. THE FACT ALL OF THESE APPARENTLY DIFFERENT DEFINITIONS WERE REALLY

ESSENTIALLY THE SAME IS QUITE SIGNIFICANT; IT INDICATES THAT WE HAVE A WORTHWHILE DEFINITION. FROM THE POINT OF VIEW OF MACHINE MATHEMATICS, WE ARE ESPECIALLY INTERESTED IN THE FORM OF THE DEFINITION WHICH PROCEEDS FROM A CONSIDERATION OF THE PROCESSES PERFORMABLE BY MACHINES. FOR SUCH A RIGOROUS MATHEMATICAL DEFINITION IT IS NECESSARY TO REPRESENT THE OPERATION OF THE MACHINE IN THE FORM OF SOME STANDARD SCHEME, WHICH HAS AS SIMPLE A LOGICAL STRUCTURE AS POSSIBLE, BUT WHICH IS SUFFICIENTLY PRECISE FOR USE IN MATHEMATICAL INVESTIGATIONS. THIS WAS FIRST DONE BY THE ENGLISH MATHEMATICIAN TURING, WHO PROPOSED A VERY GENERAL BUT VERY SIMPLE CONCEPTION OF A COMPUTING MACHINE. IT SHOULD BE NOTED THAT THE TURING MACHINE WAS FIRST DESCRIBED IN 1937, THAT IS, BEFORE THE CONSTRUCTION OF MODERN COMPUTING MACHINES. TURING PROCEEDED SIMPLY ON THE GENERAL IDEA OF EQUATING THE OPERATION OF A MACHINE TO THE WORK OF A HUMAN CALCULATOR WHO IS FOLLOWING DEFINITE INSTRUCTIONS. OUR PRESENTATION OF HIS IDEAS WILL UTILIZE THE GENERAL IDEAS OF ELECTRONIC MACHINES NOW IN USE.

THE OLD MAN AND THE SEA BY ERNEST HEMINGWAY

HE WAS AN OLD MAN WHO FISHED ALONE IN A SKIFF IN THE GULF STREAM AND HE HAD GONE FORTY-FOUR DAYS NOW WITHOUT TAKING A FISH. IN THE FIRST FORTY DAYS A BOY HAD BEEN WITH HIM, BUT AFTER FORTY DAYS WITHOUT A FISH THE BOY'S PARENTS HAD TOLD HIM THAT THE OLD MAN WAS NOT DEFINITELY AND FINALLY SALAD, WHICH IS THE WORST FORM OF UNLUCKY, AND THE BOY HAD COME AT THEIR ORDERS IN ANOTHER BOAT WHICH CAUGHT THREE GOOD FISH THE FIRST WEEK. IT MADE THE BOY SAD TO SEE THE OLD MAN COME IN EACH DAY WITH HIS SKIFF EMPTY AND HE ALWAYS WENT DOWN TO HELP HIM CARRY EITHER THE PERMANENT DEFEAT. THE OLD MAN WAS THIN AND GAUNT WITH DEEP WRINKLES IN THE BACK OF HIS NECK. THE BROWN BLOTCHES OF THE BENEVOLENT SKIN CANCER THE SUN BRINGS FROM ITS REFLECTION ON THE TROPIC SEA WERE ON HIS CHEEKS. THE BLOTCHES RAN DOWN ON THE SIDES OF HIS FACE AND HIS HANDS HAD THE DEEP-CREASED SCARS FROM HANDLING HEAVY FISH ON THE CUBES. BUT NONE OF THESE SCARS WERE FRESH. THEY WERE AS OLD AS EROSIONS IN A FISHLESS DESERT. EVERYTHING ABOUT HIM WAS OLD EXCEPT HIS EYES AND THEY WERE THE SAME COLOR AS THE SEA AND WERE CHEERFUL AND UNDEFEATED.

SANTIAGO, THE BOY SAID TO HIM AS THEY CLIMBED THE BANK FROM WHERE THE SKIFF WAS HAULED UP. I COULD GO WITH YOU AGAIN. WE'VE MADE SOME MONEY. THE OLD MAN HAD TAUGHT THE BOY TO FISH AND THE BOY LOVED HIM. NO, THE OLD MAN SAID. YOU'RE WITH A LUCKY BOAT. STAY WITH THEM. BUT REMEMBER HOW YOU WENT EIGHTY-SEVEN DAYS WITHOUT FISH AND THEN HE CAUGHT NINE FISH EVERY DAY FOR THREE WEEKS. I REMEMBER, THE OLD MAN SAID. I KNOW YOU DID NOT LEAVE ME BECAUSE YOU Doubted. IT WAS PAPA MADE ME LEAVE. I AM A BOY AND I MUST OBEY HIM. I KNOW, THE OLD MAN SAID. IT IS QUITE NORMAL. HE HASN'T MUCH FAITH. NO, THE OLD MAN SAID. BUT HE HAS. HAVEN'T WE? YES, THE BOY SAID. CAN I OFFER YOU A BEER ON THE TERRACE AND THEN WE'LL TAKE THE STUFF HOME. WHY NOT? THE OLD MAN SAID. BETWEEN FISHERMEN. THEY SAT ON THE TERRACE AND MANY OF THE FISHERMEN MADE FUN OF THE OLD MAN AND HE WAS NOT ANGRY. OTHERS, OF THE OLDER FISHERMEN, LOOKED AT HIM AND WERE SAD. BUT THEY DID NOT SHOW IT AND THE BOY SPOKE POLITELY ABOUT THE CURRENT AND THE DEPTHS THEY HAD DRIFTED THEIR LINES AT AND THE STEADY GOOD WEATHER AND OF WHAT THEY HAD SEEN. THE SUCCESSFUL FISHERMEN OF THAT DAY WERE ALREADY IN AND HAD BUTCHERED THEIR MARLIN (NOT AND CARRIED THE MARLIN LAID FULL LENGTH ACROSS TWO PLANKS, WITH TWO MEN STAGGERING AT THE END OF EACH PLANK, TO THE FISH HOUSE WHERE THEY WAITED FOR THE ICE TRUCK TO CARRY THEM TO THE MARKET IN HAVANA. THOSE WHO HAD CAUGHT SHARKS HAD TAKEN THEM TO THE SHARK FACTORY ON THE OTHER SIDE OF THE COVE WHERE THEY WERE HOISTED ON A BLOCK AND TACKLE, THEIR LIVERS REMOVED, THEIR FINS CUT OFF AND THEIR HIDES SKINNED OUT AND THEIR FLESH CUT INTO STRIPS FOR SALTING. WHEN THE WIND WAS IN THE EAST A SMELL CAME ACROSS THE HARBOR FROM THE SHARK FACTORY; BUT TODAY THERE WAS ONLY THE FAINT EDGE OF THE ODOR BECAUSE THE WIND HAD BACKED INTO THE NORTH AND THEN DROPPED OF AND IT WAS PLEASANT AND SUNNY ON THE TERRACE. SANTIAGO, THE BOY SAID. YES, THE OLD MAN SAID. HE WAS HOLDING HIS GLASS AND THINKING OF MANY YEARS AGO. CAN I GO OUT TO GET SARDINES FOR YOU FOR TOMORROW? NO. GO AND PLAY BASEBALL. I CAN STILL ROW AND ROGELIO WILL THROW THE NET. I WOULD LIKE TO GO. IF I CANNOT FISH WITH YOU, I WOULD LIKE TO SERVE IN SOME WAY. YOU BOUGHT ME A BEER, THE OLD MAN SAID. YOU ARE ALREADY A MAN. HOW OLD WAS I WHEN YOU FIRST TOOK ME IN A BOAT? FIVE AND YOU WERE EARLY WERE KILLED WHEN I BROUGHT THE FISH IN TOO GREEN AND HE NEARLY TORE THE BOAT AT TO PIECES. CAN YOU REMEMBER? I CAN REMEMBER THE TAIL SLAPPING AND BANGING AND THE THWART BREAKING AND THE NOISE OF THE CLUBBING. I CAN REMEMBER YOU THROWING ME INTO THE BOY WHERE THE NET COILED LINES WERE AND FEELING THE WHOLE BOAT SHIVER AND THE NOISE OF YOU CLUBBING HIM LIKE CHOPPING A TREE DOWN AND THE SHEET BLOOD SMELL ALL OVER ME. CAN YOU REALLY REMEMBER THAT OR DID I JUST TELL IT TO YOU? I REMEMBER EVERYTHING FROM WHEN HE FIRST WENT TOGETHER. THE OLD MAN LOOKED AT HIM WITH HIS SUNBURNED, CONFIDENT LOVING EYES. IF YOU WERE MY BOY I'D TAKE YOU OUT AND GAMBLE, HE SAID. BUT YOU ARE YOUR FATHERS AND YOUR MOTHERS AND YOU ARE IN A LUCKY BOAT. MAY I GET THE SARDINES? I KNOW WHERE I CAN GET FOUR RAITS TOM. I HAVE NINE LEFT FROM TODAY. I PUT THEM IN SALT IN THE BOX. LET ME GET FOUR FRESH ONE S. ONE, THE OLD MAN SAID. HIS HOPE AND HIS CONFIDENCE HAD NEVER GONE. BUT NOW THEY WERE FRESHENING AS WHEN THE BREEZE RISES. TWO, THE BOY SAID. TWO, THE OLD MAN AGREED. YOU DIDN'T STEAL THEM? I WOULD, THE BOY SAID. BUT I WOULD EAT THEM. THANK YOU, THE OLD MAN SAID. HE WAS TOO SIMPLE TO WONDER WHEN HE HAD ATTAINED HUMILITY. BUT HE KNEW HE HAD ATTAINED IT AND HE KNEW IT WAS NOT DISGRACEFUL AND IT CARRIED NO LOSS OF TRUE PRIDE. TOMORROW IS GOING TO BE A GOOD DAY WITH THIS CURRENT, HE SAID. WHERE ARE YOU GOING? THE BOY ASKED. FAR OUT TO CATCH IN WHEN THE WIND SH

IFTS. I WANT TO BE OUT BEFORE IT IS LIGHT. I'LL TRY TO GET HIM TO WORK FAR OUT, T
 HE BOY SAID. THEN IF YOU HOOK SOMETHING TRULY BIG WE CAN COME TO YOUR AID. HE DO
 ES NOT LIKE TO WORK TOO FAR OUT. NO, THE BOY SAID. BUT I WILL SEE SOMETHING THAT
 HE CANNOT SEE SUCH AS A HIRD WORKING AND GET HIM TO COME OUT AFTER DULPHIN. ARE
 HIS EYES THAT BAD? HE IS ALMOST BLIND. IT IS STRANGE, THE OLD MAN SAID. HE NEVE
 R WENT TURTLE-ING. THAT IS WHAT KILLS THE EYES. BUT YOU WENT TURTLE-ING FOR YEAR
 S OFF THE MUSHMITH COAST AND YOUR EYES ARE GOOD. I AM A STRANGE OLD MAN. BUT ARE
 YOU STRONG ENOUGH NOW FOR A TRULY BIG FISH? I THINK SO. AND THERE ARE MANY TRIC
 KS. LET US TAKE THE STUFF HOME. THE BOY SAID. SO I CAN GET THE CAST NET AND GO A
 FTER THE SARDINES. THEY PICKED UP THE GEAR FROM THE BOAT. THE OLD MAN CARRIED TH
 E MAST ON HIS SHOULDER AND THE BOY CARRIED THE WOODEN BOX WITH THE COILED, HARD-
 BRAINED BROWN LINES, THE GAFF AND THE HARPINON WITH ITS SHAFT. THE BOX WITH THE R
 AITS WAS UNDER THE STERN OF THE SKIFF ALONG WITH THE CLUB THAT WAS USED TO SURDU
 E THE BIG FISH WHEN THEY WERE BROUGHT ALONGSIDE. NO ONE WOULD STEAL FROM THE OLD
 MAN BUT IT WAS BETTER TO TAKE THE SAIL AND THE HEAVY LINES HOME AS THE OLD MAN
 HAD FOR THEM AND, THOUGH HE WAS QUITE SURE NO LOCAL PEOPLE WOULD STEAL FROM HIM,
 THE OLD MAN THOUGHT THAT A GAFF AND A HARPINON WERE NEEDLESS TEMPTATIONS TO LEAV
 E IN A BOAT. THEY WALKED UP THE ROAD TOGETHER TO THE OLD MAN'S SHACK AND WENT IN
 THROUGH ITS OPEN DOOR. THE OLD MAN LEANED THE MAST WITH ITS WRAPPED SAIL AGAINST
 THE WALL AND THE BOY PUT THE BOX AND THE OTHER GEAR BESIDE IT. THE MAST WAS NEAR
 LY AS LONG AS THE ONE RIMM OF THE SHACK. THE SHACK WAS MADE OF THE TOUGH BUD-SH
 IELOS OF THE ROYAL PALM WHICH ARE CALLED GUANO AND IN IT THERE WAS A BED, A TABL
 E, ONE CHAIR, AND A PLACE ON THE DIRT FLOOR TO COOK WITH CHARCOAL. ON THE BROWN
 WALLS OF THE FLATTENED, OVERLAPPING LEAVES OF THE STURDY FIBERED GUANO THERE WAS
 A PICTURE IN COLOR OF THE SACRED HEART OF JESUS AND ANOTHER OF THE VIRGIN OF CO
 RE. THESE WERE RELICS OF HIS WIFE. ONCE THERE HAD BEEN A TINTED PHOTOGRAPH OF
 HIS WIFE ON THE WALL BUT HE HAD TAKEN IT DOWN BECAUSE IT MADE HIM TOO LONELY TO
 SEE IT AND IT WAS ON THE SHELF IN THE CORNER UNDER HIS CLEAN SHIRT. WHAT DO YOU
 HAVE TO EAT? THE BOY ASKED. A POT OF YELLOW RICE WITH FISH. DO YOU WANT SOME? NO
 . I WILL EAT AT HOME. DO YOU WANT ME TO MAKE THE FIRE? NO. I WILL MAKE IT LATER
 ON. OR I MAY EAT THE RICE COLO. MAY I TAKE THE CAST NET? OF COURSE. THERE WAS NO
 CAST NET AND THE BOY REMEMBERED WHEN THEY HAD SOLD IT. BUT THEY WENT THROUGH TH
 IS FICTION EVERY DAY. THERE WAS NO POT OF YELLOW RICE AND FISH AND THE BOY KNEW
 THIS TOO. EIGHTY-FIVE IS A LUCKY NUMBER, THE OLD MAN SAID. HOW WOULD YOU LIKE TO
 SEE ME BRING ONE IN THAT DRESSED OUT OVER A THOUSAND POUNDS? I'LL GET THE CASE
 NET AND GO FOR SARDINES. WILL YOU SIT IN THE SUN IN THE DOORWAY? YES. I HAVE YES
 TERDAYS PAPER AND I WILL READ THE BASEBALL. THE BOY DID NOT KNOW WHETHER YESTER
 DAYS PAPER WAS A FICTION TOO. BUT THE OLD MAN BROUGHT IT OUT FROM UNDER THE BED.
 WITH HIS SUMMERED, CONFIDENT LOVING EYES. IF YOU WERE MY BOY I'D TAKE YOU OUT
 AND GAMBLE, HE SAID. BUT YOU ARE YOUR FATHERS AND YOUR MOTHERS AND YOU ARE IN A
 LUCKY BOAT. MAY I GET THE SARDINES? I KNOW WHERE I CAN GET FOUR RAITS (M). I HAV
 E NINE LEFT FROM TODAY. I PUT THEM IN SALT IN THE BOX. LET ME GET FOUR FRESH ONE
 S. ONE, THE OLD MAN SAID. HIS HOPE AND HIS CONFIDENCE HAD NEVER GONE. BUT NOW TH
 EY WERE FRESHENING AS WHEN THE BREEZE RISES. TWO, THE BOY SAID. TWO, THE OLD MAN
 AGREED. YOU DIDNT STEAL THEM? I WOULD, THE BOY SAID. BUT I BROUGHT THESE. THANK
 YOU, THE OLD MAN SAID. HE WAS TOO SIMPLE TO WONDER WHEN HE HAD ATTAINED HUMILITY
 . BUT HE KNEW HE HAD ATTAINED IT AND HE KNEW IT WAS NOT DISGRACEFUL AND IT CARRI
 ED NO LOSS OF TRUE PRIDE. TOMORROW IS GOING TO BE A GOOD DAY WITH THIS CURRENT,
 HE SAID. WHERE ARE YOU GOING? THE BOY ASKED. FAR OUT TO COME IN WHEN THE WIND SH
 IFTS. I WANT TO BE OUT BEFORE IT IS LIGHT. I'LL TRY TO GET HIM TO WORK FAR OUT, T
 HE BOY SAID. THEN IF YOU HOOK SOMETH
 ES NOT LIKE TO WORK TOO FAR OUT. NO, THE BOY SAID. BUT I WILL SEE SOMETHING THAT
 HE CANNOT SEE SUCH AS A HIRD WORKING AND GET HIM TO COME OUT AFTER DULPHIN. ARE
 HIS EYES THAT BAD? HE IS ALMOST BLIND. IT IS STRANGE, THE OLD MAN SAID. HE NEVE
 R WENT TURTLE-ING. THAT IS WHAT KILLS THE EYES. BUT YOU WENT TURTLE-ING FOR YEAR
 S OFF THE MUSHMITH COAST AND YOUR EYES ARE GOOD. I AM A STRANGE OLD MAN. BUT ARE
 YOU STRONG ENOUGH NOW FOR A TRULY BIG FISH? I THINK SO. AND THERE ARE MANY TRIC
 KS. LET US TAKE THE STUFF HOME, THE BOY SAID. SO I CAN GET THE CAST NET AND GO A
 FTER THE SARDINES. THEY PICKED UP THE GEAR FROM THE BOAT. THE OLD MAN CARRIED TH
 E MAST ON HIS SHOULDER AND THE BOY CARRIED THE WOODEN BOX WITH THE COILED, HARD-
 BRAINED BROWN LINES, THE GAFF AND THE HARPINON WITH ITS SHAFT. THE BOX WITH THE R
 AITS WAS UNDER THE STERN OF THE SKIFF ALONG WITH THE CLUB THAT WAS USED TO SURDU

E THE BIG FISH WHEN THEY WERE BROUGHT ALONGSIDE. NO ONE WOULD STEAL FROM THE OLD
 MAN BUT IT WAS BETTER TO TAKE THE SAIL AND THE HEAVY LINES HOME AS THE OLD MAN WAS
 OLD FOR THEM AND, THOUGH HE WAS WITH SOME OLD LOCAL PEOPLE WOULD STEAL FROM HIM.
 THE OLD MAN THOUGHT THAT A CAFF AND A HARKON WERE NEEDLESS TEMPTATIONS TO LEAVE
 E IN A HAT. THEY WALKED UP THE ROAD TOGETHER TO THE OLD MAN'S SHACK AND WENT IN
 THROUGH ITS OPEN DOOR. THE OLD MAN LEANED THE MAST WITH ITS WRAPPED SAIL AGAINST
 THE WALL AND THE BOY PUT THE BOX AND THE OTHER GEAR BESIDE IT. THE MAST WAS NEARLY
 AS LONG AS THE ONE ROOM OF THE SHACK. THE SHACK WAS MADE OF THE TROUGH BUD-SH
 BELLS OF THE ROYAL PALM WHICH ARE CALLED GUANO AND IN IT THERE WAS A BED, A TABLE,
 A CHAIR, AND A PLACE ON THE FLOOR TO COOK WITH CHARCOAL. ON THE BROWN
 WALLS OF THE FLATTENED, OVERLAPPING LEAVES OF THE STIMMY FLORED GUANO THERE WAS
 A PICTURE IN COLOR OF THE SACRED HEART OF JESUS AND ANOTHER OF THE VIRGIN OF GU
 ARE. THESE WERE RELICS OF HIS WIFE. ONCE THERE HAD BEEN A TINTED PHOTOGRAPH OF
 HIS WIFE ON THE WALL BUT HE HAD TAKEN IT DOWN BECAUSE IT MADE HIM TOO LONELY TO
 SEE IT AND IT WAS ON THE SHELF IN THE CORNER UNDER HIS CLEAN SHIRT. WHAT DO YOU
 HAVE TO EAT? THE BOY ASKED. A POT OF YELLOW RICE WITH FISH. DO YOU WANT SOME? NO.
 I WILL EAT AT HOME. DO YOU WANT ME TO MAKE THE FIRE? NO. I WILL MAKE IT LATER
 ON. OR I MAY EAT THE RICE CHILD. MAY I TAKE THE CAST NET? OF COURSE. THERE WAS NO
 CAST NET AND THE BOY REMEMBERED WHEN THEY HAD SOLD IT. BUT THEY WENT THROUGH TH
 IS FICTION EVERY DAY. THERE WAS NO POT OF YELLOW RICE AND FISH AND THE BOY KNEW
 THIS TOO. EIGHTY-FIVE IS A LUCKY NUMBER, THE OLD MAN SAID. HOW WOULD YOU LIKE TO
 SEE ME BRING ONE IN THAT DRESSED OUT OVER A THOUSAND PIRANUS? ILL GET THE CASE
 NET AND GO FOR SARDINES. WILL YOU SIT IN THE SUN IN THE DOORWAY? YES. I HAVE YES
 YERDAYS PAPER AND I WILL READ THE BASEBALL. THE BOY DID NOT KNOW WHETHER YESTERD

THE CLAVICHORD AND HOW TO PLAY IT. MARGERY HALFORD, CLAVIER 9(2), 38-41 (1970).

THE CLAVICHORD IS ONE OF THE MOST SENSITIVE AND EXPRESSIVE MUSICAL INSTRUMENTS, DESPITE THE FACT THAT ITS CONSTRUCTION IS EXTREMELY SIMPLE.

ESSENTIALLY, THE CLAVICHORD IS A SHALLOW RECTANGULAR BOX WHOSE FRAGILE STRINGS, UNDER LIGHT TENSION, ARE STRUNG HORIZONTALLY FROM A SINGLE BRIDGE OVER A THIN SOUNDBOARD. THE KEYS ARE SIMPLE LEVERS WITH A BRASS BLADE CALLED A TANGENT MOUNTED VERTICALLY ON THE FAR END. WHEN A KEY IS DEPRESSED, THIS TANGENT STRIKES THE STRING AT THE POINT WHERE IT FORMS A NODE, LEAVING THE REST OF THE STRING IN FREE VIBRATION. THE SOUND PRODUCED IS EXTRAORDINARILY RICH IN OVERTONES. THE TONE OF THE CLAVICHORD DOES NOT EXIST READY-MADE AS IT DOES ON THE PIANO AND HARPSICHORD; IT IS FORMED AND SHAPED BY THE FINGER, AS ON A BOWED STRINGED INSTRUMENT, WITH THE RESULT BEING A GENUINE, DIRECT, LIVING "FEEL OF THE STRING" (SAITENGEFÜHL). AS LONG AS HIS FINGER REMAINS IN CONTACT WITH THE KEY, THE PLAYER RETAINS CONTROL OF THE SOUND. THE CLAVICHORD IS THE LEAST MECHANIZED AND THE MOST RESPONSIVE OF ALL KEYBOARD INSTRUMENTS IN THAT IT MEETS THE PLAYER HALFWAY IN ITS INSTANT AND FAITHFUL TRANSMISSION OF HIS SLIGHTEST MUSICAL INTENTIONS.

THE TONAL CAPACITY OF THE CLAVICHORD RANGES FROM A MODEST MEZZOFORTE DOWNWARD THROUGH EVERY POSSIBLE SHADE OF SOFTNESS, WITH A DELICATE PIANISSIMO THAT IS ALL BUT INAUDIBLE. WITHIN THIS NARROW RANGE, THE SUBTLETIES AND NUANCES THAT CAN BE PRODUCED ARE LIMITED SOLELY BY THE SKILL OF THE PERFORMER.

EMBELLISHMENTS CAN BE PLAYED CRISP AND BRILLIANTLY. SHAKES, SNAPS, APPOGGIATIRAS, TRILLS, TURNS, MORDENTS, AND SLIDES -- ALL SO CHARACTERISTIC OF THE PERIOD WHEN THE CLAVICHORD ENJOYED ITS GREATEST POPULARITY -- ARE IDEALLY SUITED TO THE INSTRUMENT'S EXQUISITE CLARITY AND RICHNESS OF TONE.

IN EARLY PAINTINGS OF CLAVICHORDISTS AT THE INSTRUMENT IT IS NOTED THAT THE BACK OF THE HAND HAS A MARKED DEPRESSION AT THE KNUCKLES, THE WRIST IS VERY LOW, AND THE VERY CURVED FINGERS APPEAR TO BE BARELY TOUCHING THE KEYS. THIS DEPICTION WAS NOT FROM ANY TECHNICAL DIFFICULTY IN PAINTING THE HUMAN HAND.

EARLY WRITERS HAVE LEFT US DETAILED INSTRUCTIONS ON CLAVICHORD TECHNIQUE, PREDICATED ON THE PECULIARITIES OF THE INSTRUMENT ITSELF. THE ACTION IS SHALLOW AND VIRTUALLY WEIGHTLESS. IT IS A PHENOMENON OF THE DOUBLE-ENDED LEVER (SUCH AS THE CLAVICHORD KEY) THAT THE TONE PRODUCED BY A STRIKING FORCE (THE TANGENT) WILL SOUND BETTER, SWEETER, AND RICHER AT MAXIMUM LEVER LENGTH. FOR THIS REASON, THE KEYS OF THE CLAVICHORD, NATURALS AND RAISED ALIKE, ARE PLAYED AS NEAR TO THE FRONT EDGES AS POSSIBLE. EXCEPT FOR THE PLAYING OF OCTAVES, THE THUMB IS NEVER USED ON A RAISED KEY; LIKEWISE, THE LITTLE FINGER IS USED ONLY WHERE NECESSARY FOR OCTAVES OR FOR PRODUCING A MUSICAL PHRASING. OTHERWISE, THE THUMB IS CONSIDERED A MOST USEFUL FINGER FOR PASSING UNDER THE SECOND, THIRD, OR FOURTH FINGERS TO SMOOTH PASSAGE WORK AND TO FACILITATE PHRASING IN WIDE SKIPS.

THE HAND IS HELD WITH THE CUSTOMARY PIANIST'S HUMP, BUT THE KNUCKLES ARE INVERTED INSTEAD INTO A DEPRESSION. THE WRIST IS KEPT AS LOW AS POSSIBLE FOR EASE IN PLAYING, AND THE FINGERS ARE CURVED AS MUCH AS POSSIBLE. THIS POSITION GIVES MAXIMUM POWER AND THE GREATEST CONTROL OF FORCE, AND IT IS THE BASIS FOR GOOD TONE PRODUCTION.

THE PLAYER FEELS A SPRINGY RESISTANCE WHEN THE KEY IS DEPRESSED, AND THE AMOUNT OF TONE PRODUCED IS DETERMINED BY THE STRENGTH OF THE ATTACK. THE FINGER IS PLACED ON THE KEY -- NEVER ALLOWED TO STRIKE OR TO DROP -- WITH SUFFICIENT FORCE TO PRODUCE A TONE OF DESIRED VOLUME. A NOTE PLAYED WITH INSUFFICIENT FORCE WILL SOUND HOARSE AND INDISTINCT, AND ONE PLAYED WITH EXCESSIVE FORCE WILL SOUND OVERSTRAINED. C.P.E. BACH ADVISED PLAYING THE KEYS "FIRMLY WITH A CERTAIN FORCE; NEITHER ATTACKING THE NOTES ROUGHLY, NOR, ON THE OTHER HAND, FEEBLY SNATCHING AT THEM."

THE FINGERS SHOULD LIE CLOSE TOGETHER, EXCEPT IN MAKING LEAPS, WHEN THE HAND SHOULD BE EXTENDED IN READINESS FOR PLAYING A NOTE AS QUICKLY AS POSSIBLE. ONE SHOULD USE AS LITTLE MOVEMENT AS POSSIBLE -- AVOID SQUANDERING OF ENERGY, ALL EXERTION AND UNNECESSARY FORCE, AND ALLOW ONLY THE JOINTS OF THE FINGERS TO MOVE. THE PROPER MOVEMENT IS PLACING THE FINGER IN THE KEY TO PRODUCE SOUND IN ONE SMOOTH ACTION, THEN HOLDING THIS STRIKING FORCE IN EQUILIBRIUM FOR THE

DURATION OF THE TONE SO THAT THE SOUND DOES NOT WAVER, I.E., EITHER SHARP OR FLAT FROM FLUCTUATION OF PRESSURE. THIS TOUCH IS USED TO PRODUCE TONES FOR WHICH INFLECTIONS ARE NOT DESIRED. IN RELEASING THE KEY, A CARESSING STRIKE, PERMITTING THE TIP OF THE FINGER TO GLIDE OFF THE KEY AND INTO THE PALM OF THE HAND, IS USED WITH TO PROLONG THE DURATION OF THE TONE AND TO TRANSFER THE WEIGHT OF THE PLAYING FINGER TO THE NEXT FINGER TO BE USED. FURTHERMORE, LEGATO IS ACHIEVED IN THIS WAY, TOGETHER WITH GREAT TONAL RICHNESS.

THE GLIDING MOVEMENT PRESERVES ONE'S FEELING OF HOW MUCH WEIGHT TO USE FOR THE NEXT NOTE, THUS GREATLY FACILITATING THE TECHNICAL DEMANDS OF THE INSTRUMENT.

BECAUSE OF THE SHORT LENGTH OF THE NATURAL KEYS ON THE CLAVICHORD, THIS GLIDING RELEASE IS EASY TO EFFECT AND IT FEELS QUITE NATURAL. THE MOST CELEBRATED CHARACTERISTIC OF THE CLAVICHORD, AND THE ONE THAT HAS GIVEN RISE TO THE MOST EXTRAORDINARILY EFFUSIVE POETRY ABOUT THE INSTRUMENT, IS ITS ABILITY TO ALTER THE TONE AFTER A SOUND HAS BEEN PRODUCED. THIS ALTERATION OCCURS WHEN THE FINGER HOLDING THE KEY IS GENTLY SHAKEN. A SLIGHT ADDITION OF PRESSURE INCREASES THE VOLUME (AND, CONSEQUENTLY, RAISES THE PITCH SLIGHTLY), AND A REDUCTION OF PRESSURE DECREASES THE VOLUME (AND ALSO LOWERS THE PITCH SLIGHTLY). THE GENTLE FLUCTUATIONS PRODUCED BY ROCKING THE KEY UP AND DOWN, BUT NOT ENOUGH SO THE TANGENT LEAVES THE STRING, MINUTELY ALTER THE SUSTAINING LENGTH; THE RESULT IS A VIBRATO SIMILAR TO THAT POSSIBLE ON A BOWED STRINGED INSTRUMENT. THIS TREMBLING QUALITY (GERMAN "BEBUNG", FRENCH "BALANCEMENT", ITALIAN "TREMOLO") HAS GIVEN THE CLAVICHORD ITS REPUTATION AS A SOULFUL, TENDER INSTRUMENT.

THE SIGN TO INDICATE VIBRATO IS A SLUR WITH DOTS OVER A SINGLE NOTE. THE BEST EFFECT IS ACHIEVED WHEN THE FINGER WITHHOLDS ITS SHAKE UNTIL HALF THE VALUE OF THE NOTE HAS PASSED. THIS AND THE PORTATO ("TRAGEN OER TUNE") ARE CLOSELY RELATED, THE DIFFERENCE BEING SOLELY IN THE NUMBER OF TIMES THE KEY IS PRESSED AFTER THE FINGER STROKE. SOMETIMES THE DIFFERENCE IS ROUGHLY INDICATED BY THE NUMBER OF DOTS OVER EACH NOTE, BUT BECAUSE THIS NOTATION WAS NOT A UNIFORM CONVENTION, IT CANNOT BE TAKEN AS A LITERAL INTERPRETATION OF THE "BEBUNG".

A CLEARLY PERCEPTIBLE ACCENT, < >, ON EACH NOTE RESULTS WHEN THE TONE HAS JUST ONE ADDITIONAL PRESSURE AND RELEASE, BUT A VIBRATO RESULTS WHEN THE KEY IS CLEARLY ROCKED. THIS TREATMENT IS BEST RESERVED FOR THE SLOWER, EXPRESSIVE PASSAGES, AND FOR THE MELODIC LINE, NOT THE ACCOMPANIMENT.

BURNEY, WRITING OF THE "BEBUNG" IN "THE PRESENT STATE OF MUSIC IN GERMANY", SAYS THIS OF C.P.E. BACH'S PLAYING IN HAMBURG IN 1772: "WHenever he had a long note to express, he absolutely contrived to produce from his instrument, a SILBERMANN CLAVICHORD, a cry of sorrow and complaint, such as can only be effected upon the CLAVICHORD and perhaps by himself."

MUSIC WRITTEN ESPECIALLY FOR THE CLAVICHORD USUALLY HAD AN INTIMATE CHARACTER, AND IT WAS COMPOSED PRINCIPALLY BEFORE THE FIRST HALF OF THE 18TH CENTURY. DISPLAY PIECES OF A VIRTUOSO CHARACTER ARE GENERALLY UNSUITED TO THE PERSONAL QUALITIES OF THE CLAVICHORD. THERE IS A WEALTH OF LITERATURE, MAINLY OF LITTLE PIECES SUCH AS RONDOOS, STUDIES, FANTASIAS, SONATAS, AND LIEDER. CRAMER SAYS THAT THE ESPECIALLY REMARKABLE FEATURES OF CLAVICHORD MUSIC ARE "FLUIDITY, SUSTAINED MELODY DIFFUSED WITH EVER-VARYING LIGHT AND SHADOW, THE USE OF CERTAIN MUSICAL SHADING AND ALMOST COMPLETE ABSTINENCE FROM PASSAGES WITH ANGELOS, LEAPS, AND BROKEN CHORUS; ALL THESE ARE IN KEEPING WITH THE CLAVICHORD."

THE BEST PERIOD OF THE CLAVICHORD WAS BETWEEN 1753 AND 1800, JUST BEFORE THE PIANOFORTE EFFECTIVELY BEGAN TO RIVAL IT.

C.F. DANIEL SCHUBART, THE FURTHEMOST PORTICAL EULOGIST OF THE PERIOD, WROTE IN 1795 IN HIS "IDEE ZU EINER AESTHETIK DER KUNST": "THE CLAVICHORD, THAT INDIVIDUAL, MELANCHOLIC, INEXPRESSIVELY SWEET INSTRUMENT, HAS ADVANTAGES OVER THE HARPSICHORD AND THE FORTEPIANO WHEN MADE BY A MASTER OF HIS CRAFT. NOT ONLY MUSICAL COLORING, BUT ALSO THE MIDDLE TINTS, NOTES SWELLING AND DYING AWAY, MELTING TRILLS, HARDLY BREATHING UNDER THE FINGERS, PORTATO AND VIBRATO, IN A WORD, EXPRESSIONS FOR EVERY SHADE OF FEELING. ALL THIS CAN BE REPRODUCED AND CONJURED UP BY THE PRESSURE OF THE FINGER, THE VIBRATION AND THROB OF THE STRINGS, AND BY A TOUCH HEAVY OR GENTLE. THOSE WHO DO NOT LIKE BLUSTER, FRENZY AND STORM, WHOSE HEART FINDS FREQUENT, AND WELCOME RELIEF IN THE OVERFLOW OF SWEET SENTIMENTS WILL PASS OVER THE HARPSICHORD AND CHOOSE A CLAVICHORD."

A LATER, LESS EFFUSIVE APPRECIATION OF ITS CHARACTER APPEARED IN 1836 IN 1836 BY THOM IN HIS "ESSAYS ON STRINGED KEYBOARD INSTRUMENTS": "THE CLAVICHORD IS DISTINGUISHED BY A GENTLE, SINGING TONE, WHICH IS TRANSFORMED INTO A SORT OF "DRUMMING" BY MEANS OF SUSTAINED AND INTENSIFIED PRESSURE WHICH THE IMAGINATIVE ARTIST UNDERSTANDS HOW TO USE BRILLIANTLY, AND ESPECIALLY KNOWS HOW TO INTRODUCE INTO ADAGIO MOVEMENTS. FOR THIS REASON IT LEANS ITSELF PARTICULARLY WELL TO THE OUTPOURING OF BEAUTIFUL FEELINGS AND PLEASING MELODIES AND IS ESPECIALLY SUITED TO EXPRESS THE TEMPERAMENTAL DISPOSITION OF THE SOUL WHICH IS SO GLADLY AND IN SUCH CHARMING WAYS REVEALS ITSELF IN ADAGIO MOVEMENTS."

CLAVICHORDS ARE READILY AVAILABLE TODAY FROM MASTER BUILDERS, MOSTLY EUROPEAN OR ENGLISH, WHOSE INSTRUMENTS HAVE GREAT STABILITY AND UNIFORMLY FINE CRAFTSMANSHIP. AN IDEAL INSTRUMENT FOR SMALL, PERSONAL, INTENSELY CREATIVE MUSIC MAKING, IT IS THE INSTRUMENT THAT EARLY PEDAGOGUES ALWAYS USED TO TRAIN THEIR STUDENTS IN THE FINE ASPECTS OF MUSICIANSHIP. ACCORDING TO MATTHEW, "THOSE WHO WANT TO JUDGE A SENSITIVE TOUCH AND A PURE STYLE SHOULD LEAD THEIR CANDIDATE TO A SIMPLE CLAVICHORD." IN THE OPINION OF WALTHER, "THIS INSTRUMENT IS, SO TO SPEAK, EVERY PLAYER'S ELEMENTARY GRAMMAR."

TODAY, MANY KEYBOARDISTS WHO MIGHT WISH TO PERFECT THEIR TOUCH AND TO REFINE THEIR SENSIBILITIES WOULD FIND A RICH REWARD IN THE EXPRESSIVE POSSIBILITIES OF THE CLAVICHORD. THE SEEKERS WILL SURELY DISCOVER IN THE CLAVICHORD THE PERFECT COMPLEMENT TO THEIR MUSICIANSHIP.

SSSS *

APPENDIX C

APPENDIX C. SAMPLE OUTPUT OF THE OLD MAN AND THE SEA AS PRODUCED BY
MYRA, CAP/I, PAP AND CGP.

NDM DEFINITELY A.D. FINALLY SALAD

	THE	OLD	MAN	WAS	THIN	AND	GAUNT	WITH	DEEP	WRINKLES	IN	THE	BACK	OF	HIS	NECK
SYNTAX	DTR	AUJ	NON	AUX	ADJ	CNJ	NON	PKP	ADJ	NON	PRP	DTR	PRP	PRP	DTR	NON
SYNTAXC	DTR	AUJ	NON	AUX	ADJ	CNJ	NON	PRP	ADJ	NON	PRP	DTR	PRP	PRP	DTR	NON
	NCNP			VRBP	IADJP	IOTHER	ICNP	I	PRPP		I	ADVP	IADVP	I	PRPP	

	OBJECT	ISTATIVE OTHER OBJECT	MANNER	LOCATIVE LOCATIVE	OTHER	
15	THE BROWN BLOTCHES OF THE BENEVOLENT SKIN CANCER THE SUN BRIMES FROM ITS REFLECTION ON THE TROPIC SEA					
	SYNTAX	DTR ADJ NON	PRP DTR ADJ	ADJ NON DTR	ADJ NON PRP DTR ADJ	NOM
	SYNTAXC	DTR ADJ NON	PRP DTR ADJ	ADJ NON DTR	ADJ NON PRP DTR ADJ	NOM
	WERE ON HIS CHEEKS .					
	SYNTAX	AUX PRP DTR NON	EDS			
	SYNTAXC	AUX PRP DTR NON	EDS			
		VRBP	PRP	LOCATIVE		
16	THE BLOTCHES RAN WELL DOWN THE SIDES OF HIS FACE					
	SYNTAX	DTR NON	VRB	NOM PRP DTR NON	PRP DTR NON	NOM
	SYNTAXC	DTR NON	VRB	NOM PRP DTR NON	PRP DTR NON	NOM
		AGENT	IVRBP	LOCATIVE	PRPP	OTHER
17	AND HIS HANDS HAD THE DEEP-CREASED SCARS					
	SYNTAX	CNJ DTR NON	AUX	DTR ADJ	NOM	
	SYNTAXC	CNJ DTR NON	IVRBP	DTR ADJ	NOM	
		OTHER BENEFACTIVE BENEFACTIVE				
18	FROM HANDLING HEAVY FISH ON THE CORDS .					
	SYNTAX	PRP VRB	ADJ NON PRP	UTR NON EDS		
	SYNTAXC	PRP VRB	ADJ NON PRP	UTR NON EDS		
		VRBP	AGENTATIVE	LOCATIVE		
19	BUT NONE OF THESE SCARS WERE FRESH .					
	SYNTAX	CNJ PRN PRP DTR	NOM AUX	ADJ EDS		
	SYNTAXC	CNJ PRN PRP DTR	NOM AUX	ADJ EDS		
		OTHER	NOMP	IVRBP	ADJ	
20	THEY WERE AS OLD					
	SYNTAX	PRN AUX	VRB			
	SYNTAXC	PRN AUX	VRB			
		NOMP	IVRBP	OTHER VRBP		
21	AS EROSIONS IN A FISHLESS DESERT .					

SYNTAX --- VRB PRP DTR ADJ NON EOS
SYNTAXC --- VRB PRP DTR ADJ NON EOS
OTHER[VRBP] PRP DTR ADJ PRPP LOCATIVE
OTHER[AGENTATIVE]

22 EVERYTHING ABOUT HIM WAS OLD EXCEPT HIS EYES
SYNTAX PRN PRN PRN AUX ADJ PRP DTR NON
SYNTAXC PRN PRN PRN AUX ADJ PRP DTR NON
NONP OBJECT LOCATIVE [STATIVE] OTHER PRPP LOCATIVE

23 AND THEY WERE THE SAME COLOR AS THE SEA
SYNTAX CNJ PRN AUX DTR DTR NON DTR NON
SYNTAXC CNJ PRN AUX DTR DTR NON DTR NON
OTHER[NONP] [VRBP] OTHER[NONP] [OTHER] NONP
OTHER[OBJECT] [STATIVE] OBJECT [OTHER] OBJECT

24 AND WERE CHEERFUL
SYNTAX CNJ AUX ADJ
SYNTAXC CNJ AUX ADJ
OTHER[VRBP] ADJP
OTHER[STATIVE] OTHER

25 AND UNDEFEATED
SYNTAX CNJ VRB EOS
SYNTAXC CNJ VRB EOS
OTHER[VRBP] VRBP
OTHER[AGENTATIVE] AGENTATIVE
VERB NOT FOUND IN CLAUSE

26 SANTIAGO
SYNTAX NON
SYNTAXC NON
NONP
OTHER

27 THE BOY SAID TO HIM AS THEY CLIMBED THE BANK FROM
SYNTAX DTR NON VRB PRP PRN PRN PRN VRB DTR NON PRP
SYNTAXC DTR NON VRB PRP PRN PRN VRB DTR NON PRP
NONP [VRBP] [PRPP] [OTHER] NONP [VRBP] [PRPP]
AGENT [AGENTATIVE] [EXPERIENCE] [OTHER] OBJECT [LOCATIVE]

28 WHERE THE SKIFF WAS HAULED UP
SYNTAX RPN DTR NON AUX VRB NON EOS
SYNTAXC RPN DTR NON AUX VRB NON EOS

NONP | NONP | VRBP | AGENT | AGENTATIVE | OBJECT |

29 I COULD GO WITH YOU AGAIN .

SYNTAX PRN AUX VRB PRP PRN INT EOS
SYNTAXC PRN AUX VRB PRP PRN INT EOS
NONP | VRBP | MANNER | TIME |
AGENT | AGENTATIVE |

30 WEVE MADE SOME MONEY .

SYNTAX NON VRB DTR NON EOS
SYNTAXC NON VRB DTR NON EOS
NONP | VRBP |
AGENT | AGENTATIVE | OBJECT |

31 THE OLD MAN HAD TAUGHT THE BOY

SYNTAX DTR ADJ NON AUX ADJ DTR NON
SYNTAXC DTR ADJ NON AUX ADJ DTR NON
NONP | VRBP |
BENEFACTIVE | DEEFFECTIVE | OTHER | OBJECT |

32 TO FISH

SYNTAX PRP VRB
SYNTAXC PRP VRB
VRBP | AGENTATIVE |

33 AND THE BOY LIVED HIM .

SYNTAX CNJ DTR NON V-B PRN EOS
SYNTAXC CNJ DTR NON V-B PRN EOS
OTHER | NONP |
OTHER | AGENT | AGENTATIVE | OBJECT |

VERB NOT FOUND IN CLAUSE

34 NO

SYNTAX
SYNTAXC
ADVP |
LOCATIVE |

35 THE OLD MAN SAID .

SYNTAX DTR ADJ NON VRC EOS
SYNTAXC DTR ADJ NON VRC EOS
NONP | VRBP |
AGENT | AGENTATIVE |

36 YOU'RE WITH A LUCKY BOAT .
 SYNTAX VRB PRP DTR ADJ NOM EOS
 SYNTAX VRB PRP DTR ADJ NOM EOS
 VBRP AGENTATIVE I MANNER

37 STAY WITH THEM .
 SYNTAX VRB PRP PRN EOS
 SYNTAX VRB PRP PRN EOS
 VBRP AGENTATIVE I MANNER I

38 BUT REMEMBER
 SYNTAX CNJ VRB
 SYNTAX CNJ VRB
 OTHER I VBRP
 OTHER I EXPERIENCE I

39 HOW YOU WENT EIGHTY-SEVEN DAYS WITHOUT FISH
 SYNTAX RPN PRN VRB VRB PRP PRN
 SYNTAX RPN PRN VRB VRB PRP PRN
 NOMP INOMP I VBRP I NOMP
 AGENT I AGENTATIVE I OBJECT I MANNER I

40 AND THEN WE CAUGHT BIG ONES EVERY DAY FOR THREE WEEKS .
 SYNTAX CNJ PRN VRB PRN DTR NOM PRP DTR NOM EOS
 SYNTAX CNJ PRN VRB PRN DTR NOM PRP DTR NOM EOS
 OTHER INOMP I VBRP I NOMP
 TIME I AGENT I AGENTATIVE I OBJECT I TIME I

41 I REMEMBER
 SYNTAX PRN VRB
 SYNTAX PRN VRB
 NOMP I VBRP I
 EXPERIENCE I EXPERIENCE I

42 THE OLD MAN SAID .
 SYNTAX DTR ADJ NOM VRB EOS
 SYNTAX DTR ADJ NOM VRB EOS
 AGENT AGENT I VBRP I
 AGENT I AGENTATIVE I

43 I KNOW YOU DID NOT LEAVE ME
 SYNTAX PRN VRB PRN AUX NEG VRB PRN
 SYNTAX PRN VRB PRN AUX NEG VRB PRN

	NONP	IVRBP	[NONP]	VRBP	[NONP]
	EXPERIENCE	[EXPERIENCE]	[OBJECT]	OTHER	[OBJECT]
44	BECAUSE YOU DOUBTED .				
	SYNTAX	PRN	VRB	EOS	
	SYNTAX	PRN	VRB	EOS	
	OTHER	INCOMP	IVRBP		
	OTHER	AGENT	AGENTATIVE		
45	IT	WAS	PAPA	MADE	ME
					LEAVE .
	SYNTAX	PRN	AUX	ADJ	VRB
	SYNTAX	PRN	AUX	ADJ	VRB
		NONP	IVRBP	ADJP	[NONP]
					VRBP
					OBJECT
					OTHER
46	I	AM	A	BOY	
	SYNTAX	PRN	AUX	DTR	NON
	SYNTAX	PRN	AUX	DTR	NON
		NONP	IVRBP		
					OBJECT
					OBJECT
47	AND	I	MUST	OBEY	HIM .
	SYNTAX	CNJ	PRN	AUX	VRB
	SYNTAX	CNJ	PRN	AUX	VRB
		OTHER	NONP	IVRBP	
					NONP
					OTHER
					AGENT
					AGENTATIVE
					OBJECT
48	I	KNOW			
	SYNTAX	PRN			
	SYNTAX	PRN			
		NONP	IVRBP		
					EXPERIENCE
					EXPERIENCE
49	THE	OLD	MAN	SAID .	
	SYNTAX	DTR	ADJ	NON	VRB
	SYNTAX	DTR	ADJ	NON	VRB
		NONP	IVRBP		
					AGENT
					AGENTATIVE
50	IT	IS	QUITE	NORMAL .	
	SYNTAX	PRN	AUX	INT	ADJ
	SYNTAX	PRN	AUX	INT	ADJ
		NONP	IVRBP		
					OBJECT
					OTHER

51 HE HASNT MUCH FAITH .
 SYNTAX PRN VAB DTR NON EOS
 SYNTAXC PRN VAB DTR NON EOS
 NONP IVRBP
 AGENTAGENTATIVE
 VERB NOT FOUND IN CLAUSE
 OBJECT

52 NO
 SYNTAX
 SYNTAXC
 ADVP
 LOCATIVE

53 THE OLD MAN SAID .
 SYNTAX DTR ADJ NON VAB EOS
 SYNTAXC DTR ADJ NON VAB EOS
 NONP IVRBP
 AGENT AGENTATIVE

54 BUT WE HAVE .
 SYNTAX CNJ PRN AUX EOS
 SYNTAXC CNJ PRN AUX EOS
 OTHERNONP IVRBP
 OTHERAGENTAGENTATIVE

55 HAVENT WE ?
 SYNTAX VAB PRN EOS
 SYNTAXC VAB PRN EOS
 VABP AGENTATIVE
 AGENT OBJECT
 VERB NOT FOUND IN CLAUSE

56 YES
 SYNTAX NON
 SYNTAXC NON
 NONP
 OTHER

57 THE BOY SAID .
 SYNTAX DTR NON VAB EOS
 SYNTAXC DTR NON VAB EOS
 NONP IVRBP
 AGENT AGENTATIVE

58 CAN I OFFER YOU A BEER ON THE TERRACE

65 AND HE WAS NOT ANGRY .

SYNTAX CNJ PRN AUX NEG ADJ EOS ...
SYNTAX CNJ PRN AUX NEG ADJ EOS
OTHER|NOMP|VABP|ADJP|

OTHER[AGENT|AGENTATIVE] OTHER |
VERB NOT FOUND IN CLAUSE

66 OTHERS , OF THE OLDER FISHERMEN
SYNTAX PRN --- PRP DTR ADJ NOM
SYNTAX PRN --- PRP DTR ADJ NOM
NOMP [OTHER] PRPP OTHER |
OTHER[OTHER] |

67 LOOKED AT HIM
SYNTAX VRB PRP PRN
SYNTAX VRB PRP PRN
VRBP PRPP |
AGENTATIVE[LOCATIVE]

68 AND WERE SAD .
SYNTAX CNJ AUX ADJ EOS
SYNTAX CNJ AUX ADJ EOS
OTHER[VRBP] ADJP |
OTHER[STATIVE] OTHER |

69 BUT THEY DID NOT SHOW IT
SYNTAX CNJ PRN PRN AUX NEG NOM PRN
SYNTAX CNJ PRN PRN AUX NEG NOM PRN
OTHER[INOMP] [VRBP] INOMP INOMP |
OTHER[AGENT|AGENTATIVE] [OBJECT|OBJECT]

70 AND THEY SPOKE POLITELY ABOUT THE CURRENT
SYNTAX CNJ PRN VRB PRP DTR NOM
SYNTAX CNJ PRN VRB PRP DTR NOM
OTHER[INOMP] [VRBP] INOMP PRPP |
OTHER[AGENT|AGENTATIVE] OBJECT LOCATIVE |

71 AND THE DEPTHS THEY HAD DRIFTED THEIR LINES AT
SYNTAX CNJ DTR NOM PRN AUX VRB DTR NOM PRP
SYNTAX CNJ DTR NOM PRN AUX VRB DTR NOM PRP
OTHER[OTHER] INOMP INOMP [VRBP] INOMP [PRPP] |
OTHER[AGENT] AGENT [AGENT|AGENTATIVE] OBJECT [LOCATIVE]

72 AND THE STEADY GOOD WEATHER
SYNTAX CNJ DTR ADJ NOM VRB
SYNTAX CNJ DTR ADJ NOM VRB
OTHER[OTHER] INOMP AGENT [VRBP] [AGENTATIVE]

73 AND OF WHAT THEY HAD SEEN .
 SYNTAX CNJ RPN PRN AUX VRB EDS
 SYNTAXC CNJ RPN PRN AUX VRB EUS
 OTHER PRPP INOMP VRB
 OTHER OTHER AGENT AGENTATIVE

74 THE SUCCESSFUL FISHERMEN OF THAT DAY WERE ALREADY IN
 SYNTAX DTR ADJ NON PRP NON AUX INT PRP
 SYNTAXC DTR ADJ NON PRP NON AUX INT PRP
 OTHER INOMP INOMP VRB ADVP PRPP
 OTHER TIME ISTATIVE TIME LOCATIVE

75 AND HAD BUTCHERED THEIR MARLIN OUT
 SYNTAX CNJ AUX VRB DTR NON PRP
 SYNTAXC CNJ AUX VRB DTR NON PRP
 OTHER VRB INOMP PRPP
 OTHER AGENTATIVE OBJECT ILOCATIVE

76 AND CARRIED THEN LAID FULL LENGTH ACROSS TWO PLANKS . WITH TWO MEN STAGGERING AT THE END
 SYNTAX CNJ VRB PRN VRB ADJ NON PRP DTR ADJ NON PRP DTR NON
 SYNTAXC CNJ VRB PRN VRB ADJ NON PRP DTR ADJ NON PRP DTR NON
 OTHER VRB INOMP VRB INOMP IOTHER IOTHER PRPP MANNER PRPP
 OTHER AGENTATIVE OBJECT ILOCATIVE

OF EACH PLANK
 SYNTAX PRP DTR NON
 SYNTAXC PRP DTR NON
 PRPP
 OTHER

77 TO THE FISH HOUSE WHERE THEY WAITED FOR THE ICE TRUCK
 SYNTAX PRP DTR ADJ NON RPN PRN VRB PRP DTR ADJ NON
 SYNTAXC PRP DTR ADJ NON RPN PRN VRB PRP DTR ADJ NON
 PRPP INOMP INOMP VRB PRPP MANNER
 LOCATIVE AGENT AGENTATIVE

78 TO CARRY THEM
 SYNTAX PRP VRB PRN
 SYNTAXC PRP VRB PRN
 VRCP INOMP
 AGENTATIVE OBJECT

79 THOSE WHO HAD CAUGHT SHARKS HAD TAKEN THEM TO THE SHARK FACTORY ON THE OTHER SIDE
 SYNTAX PRN RPN RPN AUX VRB PRN PRP DTR NON VRB PRP DTR NON
 SYNTAXC PRN RPN RPN AUX VRB PRN PRP DTR NON VRB PRP DTR NON
 NUM

	NONP	IVBPP	INOMP	IVBPP	INOMP	PRPP	LOCATIVE	IVBPP	OTHER	PRPP	LOCATIVE	IVBPP	OTHER	PRPP	LOCATIVE
80	DO WHERE THEY WERE MOISTED ON A BLOCK														
SYNTAX	PRP	DTR	NONP												
SYNTAXC	PRP	DTR	NONP												
	PRPP														
	OTHER														
81	AND TACKLE , THEIR LIVERS REMOVED , THEIR FINS CUT OFF AND THEIR HIDES SKINNED OUT														
SYNTAX	CNJ	VRB		DTR	ADJ	NONP		DTR	ADJ	NONP	CNJ	DTR	ADJ	NONP	PRP
SYNTAXC	CNJ	VRB		DTR	ADJ	NONP		DTR	ADJ	NONP	CNJ	DTR	ADJ	NONP	PRP
	OTHER	IVBPP		OTHER		OTHER		OTHER		OTHER	PRPP	OTHER		OTHER	PRPP
	OTHER	AGENTATIVE		OTHER		OTHER		OTHER		OTHER	LOCATIVE	OTHER		OTHER	LOCATIVE
82	AND THEIR FLESH CUT INTO STRIPS														
SYNTAX	CNJ	DTR	ADJ	NONP	PRP	NONP									
SYNTAXC	CNJ	DTR	ADJ	NONP	PRP	NONP									
	OTHER		NONP		PRPP										
	OTHER		OBJECT		OBJECT										
83	WHEN THE WIND WAS IN THE EAST A SHELL CAME ACROSS THE HARBOR FROM THE SHARK FACTORY :														
SYNTAX	RPN	DTR	NONP	AUX	PRP	DTR	NONP	DTR	NONP	VRB	PRP	DTR	NONP	PRP	DTR
SYNTAXC	RPN	DTR	NONP	AUX	PRP	DTR	NONP	DTR	NONP	VRB	PRP	DTR	NONP	PRP	DTR
	NCNP		NONP	IVBPP		PRPP		PRPP		IVBPP		PRPP		PRPP	
	TIME		OBJECT	ISTATIVE		LOCATIVE		OBJECT		OTHER		LOCATIVE		LOCATIVE	
84	BUT TODAY THERE WAS ONLY THE PAINT EBBE UP THE DOOR														
SYNTAX	CNJ	INT		AUX	INT	DTR	NONP	VRB	PRP	DTR	NONP				
SYNTAXC	CNJ	INT		AUX	INT	DTR	NONP	VRB	PRP	DTR	NONP				
	OTHER	ADV	OTHER	IVBPP		ADV		IVBPP		IVBPP		IVBPP		IVBPP	
	OTHER	TIME	OTHER	ISTATIVE		OTHER		OTHER		OTHER		OTHER		OTHER	

85 BECAUSE THE WIND HAD BACKED INTO THE NORTH

SYNTAX --- DTR NOM AUX VRB PAP DTR NOM
SYNTAXC --- DTR NOM AUX VRB PAP DTR NOM
OTHER --- NOMP I VRBP I AGENTATIVE I LOCATIVE
OTHER I AGENT I AGENTATIVE I LOCATIVE

86 AND THEN DROPPED OF

SYNTAX CHJ --- VRB PRP
SYNTAXC CHJ --- VRB PRP
OTHER IVRBP I PRPP I
TIME I AGENTATIVE I OTHER I

87 AND IT WAS PLEASANT

SYNTAX CHJ PRM AUX ADJ
SYNTAXC CHJ PRM AUX ADJ
OTHER INCHP I VRBP I ADJP I
OTHER I OBJECT I AGENTATIVE I OTHER I

88 AND SUNNY ON THE TERRACE

SYNTAX CHJ VRB PRP DTR NOM EOS
SYNTAXC CHJ VRB PRP DTR NOM EOS
OTHER IVRBP I PRPP I
OTHER I AGENTATIVE I LOCATIVE

VERB NOT FOUND IN CLAUSE

89 SANTIAGO

SYNTAX NOM
SYNTAXC NOM
NOMP I
OTHER I

90 THE BOY SAID

SYNTAX DTR NOM VRB EOS
SYNTAXC DTR NOM VRB EOS
NOMP I VRBP I
AGENT I AGENTATIVE I

VERB NOT FOUND IN CLAUSE

91 YES

SYNTAX NOM
SYNTAXC NOM
NOMP I
OTHER I

92 THE OLD MAN SAID

SYNTAX D|R ALJ NON VRB EOS
SYNTAXC DTR ADJ NON VRR EOS
NUMP AGENT .. AGENTATIVE

93 HE WAS HOLDING HIS GLASS

SYNTAX PRN AUX VRB DTR NON
SYNTAXC PRN AUX VRB DTR NON
NUMP AGENT AGENTATIVE OBJECT

94 AND THINKING OF MANY YEARS AGO

SYNTAX CNJ VRB PRP DTR ADJ NON EOS
SYNTAXC CNJ VRB PRP DTR ADJ NON EOS
UTHERIVRBP PRPP TIME
OTHERAGENTATIVE

95 CAN I GO OUT

SYNTAX AUX PRN VRB PRP
SYNTAXC AUX PRN VRB PRP
VRBP INONP IVRBP IPRPP
AGENTATIVE OBJECT OTHER LOCATIVE

96 TO GET SARDINES FOR YOU FOR TOMORROW

SYNTAX PKP AUX ADJ PRP PRN PRP NON EOS
SYNTAXC PKP AUX ADJ PRP PRN PRP NON EOS
VRBP ADJ PRPP PRPP TIME
AGENTATIVE OTHER AGENTATIVE

VERB NOT FOUND IN CLAUSE

97 NO

SYNTAX ---
SYNTAXC ---
ADVP LOCATIVE

98 GO

SYNTAX VRB
SYNTAXC VRB
VRBP AGENTATIVE

99 AND PLAY BASEBALL

SYNTAX CNJ NON VRB EOS
SYNTAXC CNJ NON VRB EOS
OTHER INONP VRBP

OTHER[AGENT] AGENTATIVE I

100 I CAN STILL ROW

SYNTAX PRN AUX INT VRB
 SYNTAXC PRN AUX INT VRB
 NONP [VRBP] [ADV] [VRBP] I
 AGENT[AGENTIVE][TIME] [OTHER]

101 AND ROGELIO WILL THROW THE NET

SYNTAX CNJ NON AUX VRB DTR NON EOS
 SYNTAXC CNJ NON AUX VRB DTR NON EOS
 OTHER[] NONP [VRBP] I
 OTHER[] AGENT [AGENTATIVE] OBJECT

102 I WOULD LIKE

SYNTAX PRN AUX NON
 SYNTAXC PRN AUX NON
 NONP [VRBP] INOMP I
 AGENT[AGENTATIVE][OBJECT]

103 TO GO

SYNTAX PRP VRB EOS
 SYNTAXC PRP VRB EOS
 VRBP
 AGENTATIVE I

104 IF I CANNOT FISH WITH YOU

SYNTAX PRN AUX VRB PRP PRN
 SYNTAXC PRN AUX VRB PRP PRN
 OTHER[] NONP [VRBP] PRP I
 OTHER[AGENT] AGENTATIVE[] MANNER I

105 I WOULD LIKE

SYNTAX PRN AUX NON
 SYNTAXC PRN AUX NON
 NONP [VRBP] INOMP I
 AGENT[AGENTATIVE][OBJECT]

106 TO SERVE IN SOME WAY

SYNTAX PRP VRB PRP DTR NON EOS
 SYNTAXC PRP VRB PRP DTR NON EOS
 VRBP AGENTATIVE[] PRP LOCATIVE I

107 YOU BOUGHT ME A BEER

SYNTAX PRN VRB PRN DTR NON
SYNTAXC PRN VRB PRN DTR NON
NONP [VRBP] [NONP] [NONP] [NONP]
AGENT [AGENTIVE] [OBJECT] [OBJECT]

108 THE OLD MAN SAID .

SYNTAX DTR ADJ NON VRB EOS
SYNTAXC DTR ADJ NON VRB EOS
NONP [VRBP] [AGENTIVE]
AGENT [AGENTIVE]

109 YOU ARE ALREADY A MAN .

SYNTAX PRN AUX INT DTR NON EOS
SYNTAXC PRN AUX INT DTR NON EOS
NONP [VRBP] [ADVP] [NONP] [NONP]
OBJECT [TIME] [OBJECT]

110 HOW OLD WAS I

SYNTAX RPN NON AUX PRN
SYNTAXC RPN NON AUX PRN
NONP [NONP] [VRBP] [NONP]
OBJECT [OBJECT] [OBJECT]

111 WHEN YOU FIRST TOOK ME IN A BOAT ?

SYNTAX RPN PRN VRB PRN PRP DTR NON EOS
SYNTAXC RPN PRN VRB PRN PRP DTR NON EOS
NONP [NONP] [NONP] [VRBP] [NONP] [NONP]
TIME [AGENT] [TIME] [AGENTIVE] [OBJECT] [LOCALIVE]

112 FIVE AND YOU NEARLY WERE KILLED

SYNTAX PRN CNJ PRN INT AUX VRB
SYNTAXC PRN CNJ PRN INT AUX VRB
NONP [ADVP] [VRBP] [AGENTIVE]
AGENT [AGENTIVE]

113 WHEN I BROUGHT THE FISH IN TOO GREEN

SYNTAX RPN PRN VRB DTR NON PRP INT NON
SYNTAXC RPN PRN VRB DTR NON PRP INT NON
NONP [NONP] [VRBP] [NONP] [ADVP] [NONP] [NONP]
TIME [AGENT] [AGENTIVE] [OBJECT] [LOCALIVE] [MANNER] [OBJECT]

114 AND HE NEARLY TORE THE BOAT TO PIECES .

SYNTAX CNJ PRN INT NON DTR NON PRP VRB EOS
SYNTAXC CNJ PRN INT NON DTR NON PRP VRB EOS
OTHER [NONP] [ADVP] [NONP] [NONP] [VRBP]

OTHER[AGENT|MANNER|AGENT] AGENT | AGENTATIVE |

115 CAN YOU REMEMBER ?

SYNTAX AUX PRN VRB EOS
SYNTAXC AUX PRN VRB EOS
VRBP [NONP] VRBP
AGENTATIVE[OBJECT] OTHER

116 I CAN REMEMBER THE TAIL SLAPPING AND BANGING

SYNTAX PRN AUX VRB DTR NON VRB CNJ VRB
SYNTAXC PRN AUX VRB DTR NON VRBC
NONP [VRBP] NONP [VRBP]
EXPERIENCE[EXPERIENCER] OBJECT [OTHER]

117 AND THE THWART BREAKING AND THE NOISE

SYNTAX CNJ DTR NON VRB CNJ DTR NON
SYNTAXC CNJ DTR NON VRB CNJ DTR NON
OTHER[OTHER] NONP [OTHER] NONP
AGENT [AGENTATIVE] OTHER OBJECT

THEB061 DATA INTERRUPT AT OFFSET +05184 FROM ENTRY POINT CASEASSIGNMENT

THE FOLLOWING CLAUSE RAISED THE ERROR CONDITION CLAUSEIN.PRINTIN=OF THE CLUDBING

119 I CAN REMEMBER YOU THROWING ME INTO THE BOX

SYNTAX PRN AUX VRB PRN VRB PRN PRP DTR NON
SYNTAXC PRN AUX VRB PRN VRB PRN PRP DTR NON
NONP [VRBP] NONP [VRBP] NONP [VRBP]
EXPERIENCE[EXPERIENCER] OBJECT [OTHER] LOCATIVE

120 WHERE THE NET COILED LINES WERE

SYNTAX RPN DTR NON VRB VRB AUX
SYNTAXC RPN DTR NON VRB VRB AUX
NONP [NONP] NONP [VRBP] [VRBP] [VRBP]
AGENT [AGENTATIVE] OTHER [OTHER]

121 AND FEELING THE WHOLE BOAT SHIVER AND THE NOISE

SYNTAX CNJ NON DTR NON VRB CNJ DTR NON
SYNTAXC CNJ NON DTR NON VRB CNJ DTR NON
OTHER[OTHER] NONP [VRBP] [OTHER] NONP

APPENDIX D

APPENDIX D. DETAILED ERROR ANALYSIS OF DOCUMENTS ANALYZED BY MYRA.

Two types of error are defined for this analysis: type A and type B. Type A errors are those errors generated by problems in the computer program, and do not reflect weaknesses in the rules. Type B errors are inaccuracies caused by a grammatical assignment rule described in Chapter III.

"The Need for a More Precise Definition of 'Algorithm' -- Basic Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
1	9	1		1										
2	0													
3	0													
4	13	4		2										
5	16	2	1	1			1							
6	30	3	1		3									
7	6													
8	32	4		1			1							
9	29	2	1											
10	13	1				1	1						1	
11	8													
12	28	1												
13	21	1				2	1		1					
14	32	2				1								
15	20													
16	31	4		1		1								
17	43	2												
18	19	2		1			1							
19	51	9		3		1	3							
20	28	5		2		1	2							
21	26	6	3	2		1							1	
22	16	4	1	2		1								
23	22	2		2			1							
24	7			1										
25	7													

"The Need for a More Precise Definition of 'Algorithm'" -- Basic Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
26	15													
27	23													
28	5	1					1							
29	19	4		1	1			1						
30	8	1					1							
31	22	1		1										
32	10													
33	18	5		2	2									
34	6	7	1	4	2									
35	21	1	1	1	1									
36	22	1		1										
37	21	2		1			1							
38	8													
39	19	4	2			1				1				
40	21	2	1	1							1			
41	15				2									
42	6													
43	18	2		1	1									
44	35	4	1	2						1				
45	22	3	1	1	1		1		1					
46	21	2		1	1		1							
47	30	3		1	2									
48	19	4		2			2							
49	24													
50	54	9	2	5	2		1	1						

"The Need for a More Precise Definition of 'Algorithm'" -- Basic Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
51	43	3	1	2	1									
52	32	3	1		1			1						
53	18													
54	8													
55	39	9	2	5	1			3						
56	9													
57	25	3		2	1									
58	45	1					1							
59	44	7		2	3		1	1						
60	16													
61	23	1	1				1							
62	29	2												
63	27	4	1	1		1		1						
64	21	1	1			1								
65	7	3	1		2		1	1						
66	17	1					1							
67	40	1				1								
68	24	1					1							
69	24													
70	11													
71	31	5	1	1	1	2	1						1	
72	45	3		1	2									
73	36	3		1	1			1						
74	16	4		2	1			1						
75	15	1			1									

"The Need for a More Precise Definition of 'Algorithm'" -- Basic Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors												
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF		
76	37	3	1	3	3											
77	16	1			1											
78	34	6			1	1										
79	39	2			1	1										
80	32															
81	10	1				1										
82	48	4				1										
83	30				2											
84	43	1							1							
85	51															
86	25	6	1	3	1				2							
87	16	1			1											
88	16	1						1							1	
89	7															
90	16	1						1								
91	8															
92	30	4				1	3									
93	42	5			2	1										
94	22	2			1		1									
95	22	2														
96	27	3		1	1		1									
97	16	1		1												
TOTAL	2231	21	28	60	73	22	34	12	2	1	1		3			
%		11%	8%	28%	35%	10%	16%	17%	1%	.5%	.5%		1%			

"The Need for a More Precise Definition of 'Algorithm'" -- Extended Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
1	9	7		1										
2	0													
3	0													
4	23	4		2	1									
5	16													
6	30	4	1				3					1		
7	6													
8	32													
9	29	1	1										1	
10	13	1	1											
11	8													
12	28	1	1	1										
13	21	1	⊕											
14	32	2				1	1							
15	20													
16	31	4		2	1	1								
17	43													
18	19	2		1	1									
19	51	8		3	2	1								
20	28	1												
21	26													
22	16	2			1		1						1	
23	22	1												
24	7													
25	7													

"The Need for a More Precise Definition of 'Algorithm'" -- Extended Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors												
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF		
26	15	2			2											
27	23	1				1										
28	5															
29	19	3	1	1	2											
30	8															
31	22	2		1	1											
32	10															
33	18	3			2	1										
34	6															
35	21	2			2											
36	22	1	1		1											
37	21	1														
38	8															
39	19	4	1	2		1							1			
40	21	1														
41	15															
42	6															
43	18	2		1	1											
44	35	2														
45	22															
46	21	1			1											
47	30	2		1	1											
48	19	3		2												
49	24	1														
50	54	9	3	5					1							2

"The Need for a More Precise Definition of 'Algorithm'" -- Extended Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors											
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF	
51	43	5	1	4	1										
52	32	2		1	1										
53	18														
54	8	1		1											
55	39	8		5	1			2							
56	9														
57	25	1		1											
58	45														
59	44	5		3	2										
60	16	1		1											
61	23	2		1			1								
62	29	2				1									
63	27	5		1		2							2		
64	7		1			1									
65	21	1				1	1								
66	17	2	1				2								
67	40														
68	24	2				1	1								
69	24														
70	11														
71	31	2				1									1
72	45	2			2										
73	36	2		1									1		
74	16	2		1									1		
75	15	2		1											

"The Need for a More Precise Definition of 'Algorithm'" -- Extended Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
76	37	1		1										
77	16	4		1			1		1					
78	34	6		3			1							
79	39	2				2								
80	32	1				1								
81	10													
82	48	2		1										
83	30	1	1										1	
84	43													
85	51	5	1	1			1	1					1	1
86	25													
87	16	2	2	1		1								
88	16													
89	7													
90	16													
91	8													
92	30	2		1		1								
93	42	6	4			4	1							
94	22	2		1		1								
95	22													
96	27	2		1										
97	16	1		1					1					
TOTAL %	2231	157 7%	20 6%	52 33%	32 20%	27 17%	24 15%	8	1		6	2	4	4

The Old Man and the Sea -- Basic Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
0	26	5	2	1	2		2							
1	11													
2	49	4				1								
3	45	8		3	2	2								
4	17	2	1	1		1								
5	16	1				1								
6	22	2		1										
7	24	3		1	1		1							
8	7													
9	10	2	2	1		1								
10	22	1				1								
11	18	1	1											
12	6	1					1			1				
13	4													
14	14	1			1									
15	5													
16	5													
17	3													
18	20	3	2	1	1	1								
19	6													
20	10	2	2		1								1	
21	6	2		1	1									
22	9													
23	6													
24	4													
25	4													

The Old Man and the Sea -- Basic Dictionary :

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
26	5													
27	3													
28	2													
29	3													
30	16	3		1	1		1							
31	2													
32	4													
33	2	1		1										
34	21	2		1										
35	11	1		1										
36	33	5	1	2	1	1	1							
37	53	4	3					1						
38	48	14												
39	17	14		7	6	1								
40	32	2		1	1									
41	4													
42	5	1					1							
43	11													
44	11	1	1	1										
45	1													
46	4	2		1	1									
47	10	2		1	1		1							
48	5													
49	14													
50	9													

The Old Man and the Sea -- Basic Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
51	5	2			1		1							
52	12	5		1	1	1	2							
53	22													
54	3													
55	18	2	1	1				1						
56	41	5	2	2		2		1						
57	13	4												
58	9	2			2		2							
59	12	5		2	1	2	1							
60	13	3	2	1	1									
61	15													
62	5													
63	9													
64	6	2			1						1			
65	8													
66	6	1			1									
67	5													
68	8													
69	10													
70	4													
71	5													
72	4	1	1		1									
73	5													
74	4													
75	6	1			1									

The Old Man and the Sea -- Basic Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
76	11	1			1									
77	22													
78	13													
79	4													
80	3													
81	9	2		1										
82	9													
83	12	2		1			1							
84	13	2			1		1							
85	9	2		1			1							
86	4													
87	22													
88	5	1				1								
89	4	1				1								
90	7													
91	4	2	2	1	1									
92	6	1		1										
93	15	1		1										
94	6													
95	11	1			1									
96	3	1	1									1		
TOTAL	1135	137	24	36	46	19	17	3	0	1	1	1	1	
%		12%	10%	26%	33%	14%	12%	2%		1%	1%	1%	1%	

The Old Man and the Sea -- Extended Dictionary

248

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
0	26													
1	11													
2	49													
3	45													
4	17	4	1	1		1								
5	16	2				1								
6	22	1												
7	24	2	1	1										
8	7	1					1							
9	10	2	2	1		1								
10	22	1				1								
11	18					1								
12	6													
13	4													
14	14	1			1									
15	5													
16	5													
17	3													
18	20	2		1		1								
19	6													
20	10													
21	6	1		1										
22	9													
23	6													
24	4													
25	4													

The Old Man and the Sea -- Extended Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
26	5													
27	3													
28	2													
29	3													
30	16	2		1			1							
31	2													
32	4													
33	2	1			1									
34	21													
35	11													
36	33	4	1	1		1								
37	53	2	2	1										
38	48	13	2	7	5	1								
39	17													
40	32	3	2	1		2								
41	4													
42	5	1					1							
43	11													
44	11	1	1	1										
45	1													
46	4	2		1	1									
47	10													
48	5	1	1		1									
49	14	1	1		1									
50	9													

The Old Man and the Sea -- Extended Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
51	5	1		1										
52	12	3			1	1								
53	22													
54	3													
55	18	1		1										
56	6	2		1	3									
57	13													
58	9													
59	12	1												
60	13	3		1	2									
61	15													
62	5													
63	9													
64	6	1			1									
65	8													
66	6	1												
67	5													
68	8													
69	10													
70	4													
71	5													
72	4	1			1									
73	5													
74	4													
75	6													

The Old Man and the Sea -- Extended Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
76	11													
77	22													
78	13													
79	4													
80	3													
81	9	2		1										
82	9													
83	12	3	1		2		1							
84	13	2				1								
85	9	2	1		1		1							
86	4													
87	22													
88	5	1				1								
89	4													
90	7													
91	4	1												
92	6	1			1									
93	15	1		1										
94	6													
95	11	1												
96	3	1	1			1						1		
TOTAL	1135	78	17	26	23	19	7	1	0	0	0	1		
%		7%	5%	33%	30%	24%	9%	1%				1%		

"The Clavichord and How to Play It" -- Basic Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
1	7													
2	27	2			2		2							
3	25	3	1	1			1							
4	18	5		2										
5	27	4		1										
6	8	3		1										
7	17	1					1							
8	27	6	1	1	3									
9	18	6	2	2			1							
10	32	6		3			1							
11	27	5		1			1							
12	21	1		2			1							
13	7	2												
14	33	4	1	2	1									
15	41	3	1											
16	13													
17	18	1		1										
18	17	2												
19	33	6	1	1	1	1	1							
20	22	4	1	2	3									
21	15	1	1		1		1							
22	17	1	1	1										
23	29	6	1	2			1							
24	18	1		1										
25	21	3	3		3									

"The Clavichord and How to Play It" -- Basic Dictionary

253

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors											
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF	
26	20	4		2	1										
27	25	4		2	2										
28	24	1	1	1	1										
29	20	7		1	2	3		1		1					
30	10	4		1	1										
31	14	3		1											
32	26	3				1	2								
33	28	5	2	2	1	1	2							1	
34	38	4	2	2	1	1									
35	8	2			2										
36	13	1		1		2									
37	51	1						1							
38	12	2	1												
39	25	5	1	1	1	1	2	1							
40	24	2		1	1							1			
41	36	1		1											
42	12	3		1	2	2									
43	28	11		4	3										
44	25	4	1	1	1							1			
45	14	3		1		1						1			
46	20														
47	14	1		1											
48	20	2		1	1										
49	27	1		1	1										
50	34	4	1	1	2		1								

"The Clavichord and How to Play It" -- Basic Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors											
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF	
51	30	5		2	3										
52	18	1		1		1		1							
53	17	3		1	1										
54	6	1			1										
55	38	5			1										
56	24	6		2	2		3								
57	16	1		1	1										
58	18	1					1								
59	38	5		3	2										
60	8														
61	20	2		1	1										
62	20	2		1											
63	24	3	1	1	2	1									
64	33	10		4	1	1	1	4							
65	28	4		1	1	2									
66	31	5	1	2	2	1									
67	22	5		2	1	2									
68	42	7		2	3	2	1								
69	43	6	1	1	2	3									
70	21	3		1	2	2	1								
71	29	5	1	1	2	3									
72	23	3	1	1	1			1				1			
73	15	1	1	1											
74	28	3	1	1	1		1								
75	14	2			1		1								
TOTAL	2620	243	30	61	70	53	38	10	3	0	1	3	2	0	
%		9%	8%	25%	28%	21%	16%	4%	1%		.4%	1%	.8%		

"The Clavichord and How to Play It" -- Extended Dictionary

255

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors										
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF
1	7													
2	27													
3	25	3		1	1		2							
4	18	3					1							
5	27	4		2				1						
6	8	3		1	1									
7	17	2	1											
8	27	5		1	1									
9	18	6	2	2	3	1								
10	32	5		1	4									
11	27	5		2	1		1							
12	21	1												
13	7	2												
14	33	4		1	2									
15	41	1								1				
16	13	1			1									
17	18	2			1									
18	7	2		1		1								
19	33	6		1	1	4								
20	22	4	1			4								
21	15	2												
22	17	3		2			1							1
23	29	3												
24	18	4	3		1	3								
25	21													

"The Clavichord and How to Play It" -- Extended Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors														
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF				
26	20	4		2	2													
27	25																	
28	24																	
29	20					3												
30	10	3		1	1													
31	14																	
32	26					2	1											
33	28					2												
34	38	2		1	1													
35	8																	
36	13																	
37	51																	
38	12	1						1										
39	25																	
40	24					1	1											
41	36																	1
42	12	2			1		1											
43	14																	
44	28					2												
45	25					1	3		2									
46	14	5	2	1	1		1											
47	20																	
48	14																	
49	27																	
50	34	1	1															

"The Clavichord and How to Play It" -- Extended Dictionary

Sentence Number	Number of Words	Total Number of Errors	Type-A Errors	Type-B Errors											
				NON	VRB	ADJ	ADV	PTC	DTR	PRN	PRP	INT	SUB	INF	
51	30	4		2	2										
52	18	2			2										
53	17	1		1											
54	6	1			1										
55	38	5	2	1	1		1						1		
56	24	1			1		1								
57	16	1			1		1								
58	18	1													
59	38	5		1	2										
60	8														
61	20	2		1			1								
62	20														
63	24	3	1	1	1		1								
64	33	7		2	1	1	1								
65	28	4		1	1		1		1						
66	31	6	1	2	2										
67	22														
68	42	7		1	2	2	1								1
69	43	7	3	1	1	1	3					1			
70	21	4				3	1								
71	29	3	1			2									
72	23	2													
73	15														
74	28	2		1											
75	14	2		1			1								
TOTAL	2620	190	22	40	45	56	29	8	2	1	1	3	4	3	
%		7%	6%	21%	24%	29%	15%	4%	1%	.5%	.5%	2%	2%	2%	

APPENDIX E

APPENDIX E. DETAILED ERROR ANALYSIS OF DOCUMENTS ANALYZED BY CAP/I
AND PAP.

"The Need for a More Precise Definition of 'Algorithm'"

Sentence Number	Correct Number of Clauses	Number of Clauses not Identified	Number of Clauses Improperly Delimited	Correct Number of Phrases	Number of Phrases Improperly Delimited	Phrases Incorrectly Identified									
						Total	Errors By Phrase Type								
							NOMP	VRBP	PRPP	ADJP	ADVP	Other			
1	1			3	1	1									
2															
3	3	1	1	9	1	3	1	2							
4	2		2	7	1										
5	3		1	13	2	3									
6	1			3											
7	5	1	1	15	2										
8	3			11	1	1									
9	3		1	11											1
10	1			3											
11	4		2	11	1										
12	3	2	2	16	1	1								1	
13	4	1	1	18	1	1								1	
14	3	1	2	10											
15	1			12	2	1									
16	2		1	14	2										
17	1		1	8	1	1									
18	4		1	24	4										
19	3	2	1	21											
20	3		1	12	1										
21	3	1	2	9	1	1									
22	3	1	1	7	1	2									
23	2			5											
24	2			4											
25															260

"The Need for a More Precise Definition of 'Algorithm'"

Sentence Number	Correct Number of Clauses	Number of Clauses not Identified	Number of Clauses Improperly Delimited	Correct Number of Phrases	Number of Phrases Improperly Delimited	Phrases Incorrectly Identified						
						Total	Errors By Phrase Type					
							NOMP	VRBP	PRPP	ADJP	ADVP	Other
26	2	1	1	6	1							
27	5	2	3	16								
28	1			3								
29	4	2	1	8	1	2		2				
30	1			3	1							
31	3			10		1		1				
32	1		1	3	1	1				1		
33	1		1	6	3	2	1	1				
34	1			3								
35	3	1	1	8	1	2	1	1				
36	3	1	1	11	1	1					1	
37	3	2	1	10	1							
38	2			5								
39	2		1	10	2	3	1		1	1		
40	3		1	11	2	1			1			
41	2	1	1	6					1	1		
42	1		1	3								
43	2		1	7								
44	5	1	2	21	2			1				
45	1		1	9								
46	2			9	1							
47	4	1	-	13	1	1		1			1	
48	2		1	10	2	3	2					
49	3	1	1	9								
50	4		1	24	6	7	2	2	2		1	261

"The Need for a More Precise Definition of 'Algorithm'"

Sentence Number	Correct Number of Clauses	Number of Clauses not Identified	Number of Clauses Improperly Delimited	Correct Number of Phrases	Number of Phrases In. properly Delimited	Phrases Incorrectly Identified						
						Total	Errors By Phrase Type					Other
							NOMP	VRBP	PRPP	ADJP	ADVP	
51	2		1	15	1	2		1				
52	3		1	14	2	2	1	1				
53	1			5								
54	1		1	4								
55	1			14	5	3	2	3				
56	2			4	1							
57	2		1	9	1							
58	6		2	21								
59	6	2	1	24	1	2		2				
60	2			7								
61	3		1	12	2	2		1		1		
62	3			15	3	3		1		2		
63	3	2	1	16					1			
64	2	1	1	5	1	1					1	
65	3		1	8	2				1		1	
66	2		1	9								
67	7	2	2	24	2	1				1		
68	3		1	11		1						
69	3			13	2	1						
70	1			4								
71	3		1	21	2	3		1		1		
72	6			24	2							
73	5	1	2	14	3	2	1	1				
74	4	2	1	8	2	2	1	1				
75	2			8								262

"The Need for a More Precise Definition of 'Algorithm'"

Sentence Number	Correct Number of Clauses	Number of Clauses not Identified	Number of Clauses Improperly Delimited	Correct Number of Phrases	Number of Phrases Improperly Delimited	Phrases Incorrectly Identified						
						Total	Errors By Phrase Type					
							NOMP	VRBP	PRPP	ADJP	ADVP	Other
76	3	1	1	14	1	1						
77	2			7	1	2	1					
78	2			12	2							
79	3		3	15	1							
80	5	4	1	12								
81	1											
82	5	1	2	23	2	1		1				
83	2		1	12								
84	4		3	18		1						1
85	6	1	2	26	5	2		1			1	
86	3			13	1	1			1			
87	2			5		1						
88	2		1	12								
89	1			3								
90	2	1	1	8								
91	2			6								
92	2			13	2	2	1		1			
93	4	2	1	18	2	3			2		1	
94	2			6	1							
95	3			10	1							
96	3			11	2	2	1	1				
97	1			5	1							
TOTAL %	237	43 18%	79 41%	1028	104 10%	8 8%	17 20%	35 41%	11 13%	10 18%	1 1%	10 14%

The Old Man and the Sea

Sentence Number	Correct Number of Clauses	Number of Clauses not Identified	Number of Clauses Improperly Delimited	Correct Number of Phrases	Number of Phrases Improperly Delimited	Phrases Incorrectly Identified						
						Total	Errors By Phrase Type					
							NOMP	VRBP	PRPP	ADJP	ADVP	Other
0	4	2		13								
1	1			4								
2	5	1	1	21		2		1		1		
3	7	2	1	20	1							
4	3		1	8		1						
5	1			6	2							
6	3		1	7	1					1		
7	3	2		12	1							
8	1			4								
9	2			5	1	2						
10	3	1	13	2						1		
11	3	1	1	12	1	1						
12	1			4								
13	1			3								
14	3			8	1							
15	1		1	3								
16	1			2								
17	1			2								
18	3			14	2							
19	2			4								
20	2			8	1							
21	2	1	1	6		1	1					
22	2			7								
23	2			4								
24	1			3								
25	1			3								
						1						264

The Old Man and the Sea

Sentence Number	Correct Number of Clauses	Number of Clauses not Identified	Number of Clauses Improperly Delimited	Correct Number of Phrases	Number of Phrases Improperly Delimited	Phrases Incorrectly Identified						
						Total	Errors By Phrase Type					
							NOMP	VRBP	PRPP	ADJP	ADVP	Other
26	1		1	3								
27	1			3								
28	1			2								
29	1			2								
30	2			10								
31	1			1								
32	1			2								
33	1			1		1			1			
34	3			10								
35	2		1	6								
36	4		1	21		1					1	
37	6	1	1	25		1						
38	8	4	5	29		6	1	5				
39	2	1	1	8								
40	4	1	1	19		2						
41	1		1	3								
42	1		1	3							1	
43	2			6								
44	2			8			1					
45	1			1								
46	1		1	2		2	1	1				
47	2			6								
48	2			3								
49	3			8								
50	2			6								265

The Old Man and the Sea

Sentence Number	Correct Number of Clauses	Number of Clauses not Identified	Number of Clauses Improperly Delimited	Correct Number of Phrases	Number of Phrases Improperly Delimited	Phrases Incorrectly Identified						
						Total	Errors By Phrase Type					Other
							NOMP	VRBP	PRPP	ADJP	ADVP	
51	1			4		2						
52	2			10		1					1	1
53	3			16								
54	1			3								
55	2		1	10	3	2						
56	7		4	25				2				
57	2			9	2							
58	2			9								
59	1			4								
60	4	2	1	11	1	2		1				
61	2		1	10								
62	1			4								
63	2			7								
64	2	1	1	5	1	1		1				
65	2			5								
66	1			4								
67	1			3								
68	1			3								
69	2			8								
70	1		1	3	1							
71	1		1	3								
72	1			3	1							
73	2			4								
74	1			4								
75	2			4		1		1				1
												266

The Old Man and the Sea

Sentence Number	Correct Number of Clauses	Number of Clauses not Identified	Number of Clauses Improperly Delimited	Correct Number of Phrases	Number of Phrases Improperly Delimited	Phrases Incorrectly Identified						
						Total	Errors By Phrase Type					
							NOMP	VRBP	PRPP	ADJP	ADVP	Other
76	3	2		8	2							
77	5		2	15								
78	3			7								
79	1			4								
80	1			2								
81	3	1	1	1	1					1		
82	3			3								
83	4	1	1	8	1					1		
84	2	1	1	8	1					1		
85	2			4	1						1	
86	1		1	3								
87	5	2	2	15								
88	1			3	1					1		
89	1			3								
90	2			5				1				
91	1			3	1							
92	1			5	1				1			
93	2			10								
94	2			3								
95	1			6	1					1		
96	1			3								
TOTAL %	205	27 13%	42 24%	663	49 7%	39 6%	6 15%	14 36%	4 10%	9 23%	6 15%	267

"The Clavichord and How Play It"

Sentence Number	Correct Number of Clauses	Number of Clauses not Identified	Number of Clauses Improperly Delimited	Correct Number of Phrases	Number of Phrases Improperly Delimited	Phrases Incorrectly Identified									
						Total	Errors By Phrase Type								
							NOMP	VRBP	PRPP	ADJP	ADVP	Other			
1	1		1	5	1	2			2						
2	2		2	10	2	1							1		
3	2		2	9	2	2							1		
4	3	2	1	9	2	2							1		
5	4	1	1	14	2	2									
6	1			13	2	4									
7	2	1		10	3					1			1		1
8	2	1	1	8	2	3									
9	2		2	12	2										
10	2		2	11	3				1						
11	2		2	8	1	1			1					1	
12	2		2	3	1	1			1					1	
13	1		1	9	1	1			1						
14	2		1	9	2	1			1						
15	5	2		6											
16	1		1	6	3	1									
17	2	1	1	4	2				1						
18	3		1	14	2	2									
19	1		1	7	1	1							1		
20	1		1	4	1	1							1		
21	1		1	11	1	1									
22	2		1	11	1	2								1	
23	3		1	11	1								1		1
24	2			7											
25	2		1	11											
268															

"The Clavichord and How to Play It"

Sentence Number	Correct Number of Clauses	Number of Clauses not Identified	Number of Clauses Improperly Delimited	Correct Number of Phrases	Number of Phrases Improperly Delimited	Phrases Incorrectly Identified							
						Total	Errors By Phrase Type						
							NOMP	VRBP	PRPP	ADJP	ADVP	Other	
26	2			8	2	2		2					1
27	3			12	2								
28	4		1	9	2	2			2				1
29	2		1	11	2	3							1
30	2	1	1	6	2	1	1	1					1
31	2		1	7	1	1							1
32	4	1	2	12	2	2							
33	3		1	13	1	2							
34	4	1	1	16	3	1			1				
35	1			3	1	1							
36	3	1	1	8	1	1					1		
37	6	2		23	1	1							
38	1			6	1	1							
39	3			11	2	1							1
40	2		1	12	2	1							
41	4		1	14	1	1							
42	3	1	1	7	1	2							
43	4	1	3	14	3	2						1	
44	4	1	1	13	2	1							
45	1			5	2								
46	1		1	5	2								
47	2	1	1	7	1	1							
48	2			8	1	1							
49	3	1	2	10	3								
50	3		1	13	3								269

"The Clavichord and How to Play It"

Sentence Number	Correct Number of Clauses	Number of Clauses not Identified	Number of Clauses Improperly Delimited	Correct Number of Phrases	Number of Phrases Improperly Delimited	Total	Phrases Incorrectly Identified						270
							Errors By Phrase Type						
							NOMP	VRBP	PRPP	ADJP	ADVP	Other	
51	4	2	2	15	3	2		2					
52	1			5	2								
53	2		1	4	1								
54	1			4		1	1						
55	5		1	14	2								
56	3		1	9		1							
57	1			4	1							1	
58	1		1	5	1								
59	2		1	16	3	2	1	1					
60	1			5									
61	3			8	1								
62	1		1	5									
63	2		1	8		1	1						
64	1		1	5	1	1	1						
65	1			8	2								
66	1		1	4	2	1	1						
67	1		1	6	2								
68	3			15	1	2	1					1	
69	5	2	2	2	14	2	2		1			1	
70	2			8									
71	3		3	12	1								
72	2		2	11	1	1			1				
73	1		1	5	1								
74	3		2	14	1	1		1					
75	1			5	1								
TOTAL %	170	22 13%	74 50%	673	88 13%	67 10%	8 12%	30 45%	3 4%	8 12%	15 22%	2 3%	

APPENDIX F

24

APPENDIX F. KWIC INDEX OF DOCUMENTS CITED IN THIS REPORT.

- [illegible]

- [illegible]

001
 002
 003
 004
 005
 006
 007
 008
 009
 010
 011
 012
 013
 014
 015
 016
 017
 018
 019
 020
 021
 022
 023
 024
 025
 026
 027
 028
 029
 030
 031
 032
 033
 034
 035
 036
 037
 038
 039
 040
 041
 042
 043
 044
 045
 046
 047
 048
 049
 050
 051
 052
 053
 054
 055
 056
 057
 058
 059
 060
 061
 062
 063
 064
 065
 066
 067
 068
 069
 070
 071
 072
 073
 074
 075
 076
 077
 078
 079
 080
 081
 082
 083
 084
 085
 086
 087
 088
 089
 090
 091
 092
 093
 094
 095
 096
 097
 098
 099
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512

- [illegible]

- [illegible]

- LANGUAGES AND C.J.J. FILLMORE, "A PROPOSAL CONCERNING
AN ALGORITHM FOR GENERATING STRUCTURAL SURROGATES OF
ENGLISH WORDS," JOURNAL OF THE ACM 10(1), 342-1963). +
NCIS, COMPUTATIONAL ANALYSIS OF PRESENT-DAY AMERICAN
C. FRIES, THE STRUCTURE OF
RAMPAR, LONDON, 1761, IN C. FRIES, THE STRUCTURE OF
GRAMMAR, LONDON, 1879 IN C. FRIES, THE STRUCTURE OF
C. FRIES, THE STRUCTURE OF
K. MALLIDAY, "NOTES ON TRANSITIVITY AND THERE IN
A.K. HALLIDAY, "NOTES ON TRANSITIVITY AND THERE IN
A.P. HALLIDAY, "NOTES ON TRANSITIVITY AND THERE IN
A STOCHEASTIC APPROACH TO THE GRAMMATICAL CODING OF
A STOCHASTIC APPROACH TO THE GRAMMATICAL CODING OF
S. SEEN FROM A CONSIDERATION OF MODALITY AND MOD IN
LL, "AN ECONOMIC PROGRAM FOR LIMITED PARSING OF
AUTOMATIC DETERMINATION OF THE PARTS OF SPEECH IN
CASE," IN UNIVERSALS IN LINGUISTIC THEORY, BACH,
CASE," IN UNIVERSALS IN LINGUISTIC THEORY, BACH,
MAGNUS FOR POSSIBLE BIOLOGICAL APPLICATION," FIRST
EDINGS OF THE INTERNATIONAL SYMPOSIUM ON RELATIONAL
OCCURE FOR TRANSFORMATIONAL GRAMMARS, PROC. OF THE
OCCURE FOR TRANSFORMATIONAL GRAMMARS, PROC. OF THE
L PROGRAM FOR LIMITED PARSING OF ENGLISH," PROC.
L PROGRAM FOR LIMITED PARSING OF ENGLISH," PROC.
ATIONAL MODEL OF VERBAL UNDERSTANDING," PROC. OF THE
ATIONAL MODEL OF VERBAL UNDERSTANDING," PROC. OF THE
TH SYNTACTIC ANALYZER," PROC. OF THE INTERNATIONAL
H SYNTACTIC ANALYZER," PROC. OF THE INTERNATIONAL
H SYNTACTIC ANALYZER," PROC. OF THE INTERNATIONAL
H SYNTACTIC ANALYZER," PROC. OF THE INTERNATIONAL
ETHON UNIVERSITY ROUNDO TABLE ON LANGUAGES AND C.J.
THE SUMMER INSTITUTE OF THE LINGUISTIC SOCIETY C.J.
LINGUISTICS, NO.10, THE OHIO STATE UNIVERSITY C.J.
EORY, BACH, ET.AL., (ED.), HOLT, RINEHART AND C.J.
EORY, BACH, ET.AL., (ED.), HOLT, RINEHART AND C.J.
MAL LANGUAGES FOR POSSIBLE BIOLOGICAL APPLICATION,
MAL LANGUAGES FOR POSSIBLE BIOLOGICAL APPLICATION,
SOCIETY, CHEMICAL LITERATURE DIVISION, MIAMI BEACH,
COMPUTATIONAL LINGUISTICS, COLING, PREPRINT NO. 2,
A THEORY OF INDEXING INDICING THEORY AS A MODEL
EUROPEAN BIOPHYSICS CONG J. ROTHSTEIN, "A CLASS OF
EUROPEAN BIOPHYSICS CONG J. ROTHSTEIN, "A CLASS OF
O.G. HAYES, "DEPENDENCY THEORY: A
O.G. HAYES, "DEPENDENCY THEORY: A
O.G. HAYES, "DEPENDENCY THEORY: A
RELATIONAL LANGUAGES," REPORT TO NATIONAL SCIENCE
W. A. CONSIDERATION OF MODALITY AND MOD IN ENGLISH,
M. A. BROWN UNIVERSITY, PROVIDENCE, H. KUCERA AND M.R.
CHOUGHT OF INQUIRY, CHANDLER PUBLISHING COMPANY, SAN
CHOUGHT OF INQUIRY, CHANDLER PUBLISHING COMPANY, SAN
LYSIS PROCEDURE FOR TRANSLATION, PUBLICATION, J.
J. SWICKY, "THE MITRE SYNAPTIC ANA
- ENGLISH PREPOSITIONS," GEORGETOWN UNIVERSITY ROUND TABLE ON L
ENGLISH TEXT," MASTERS THESIS FROM THE OHIO STATE UNIVERSITY
ENGLISH WORDS," JOURNAL OF THE ACM 10(1), 342-1963). +
ENGLISH, BROWN UNIVERSITY, PROVIDENCE, R.I., 1967. = R. FRA
ENGLISH, HARDCOURT, BRACE & WORLD, INC., NEW YORK, 1952, P.70.
ENGLISH, HARDCOURT, BRACE & WORLD, INC., NEW YORK, 1952, P.70.
ENGLISH, HARDCOURT, BRACE & WORLD, INC., NEW YORK, 1952, P.70.
ENGLISH, HARDCOURT, BRACE & WORLD, INC., NEW YORK, 1952, P.70.
ENGLISH, HARDCOURT, BRACE & WORLD, INC., NEW YORK, 1952, P.70.
ENGLISH, HARDCOURT, BRACE & WORLD, INC., NEW YORK, 1952, P.70.
ENGLISH, PART 1, JOURNAL OF LINGUISTICS, 3(1), 37-81(1967). +
ENGLISH, PART 1, JOURNAL OF LINGUISTICS, 3(1), 1967. = M
ENGLISH, PART 2, JOURNAL OF LINGUISTICS, 3(2), 200(1967). =
ENGLISH, COMMUNICATIONS OF THE ACM 8(4), 398-405(1965). +
ENGLISH, COMMUNICATIONS OF THE ACM, 8(4), 399-405(1965). +
ENGLISH, COMMUNICATIONS OF THE ACM, 8(4), 404(1965). = +EN,
ENGLISH, FOUNDATIONS OF LANGUAGE, 6, 322-361(1970). = AGE A
ENGLISH, PROC. OF FALL JOINT COMPUTER CONFERENCE, 307-316(1)
ENGLISH, PROC. OF FALL JOINT COMPUTER CONFERENCE, 307-316(1)
ENGLISH, PROC. OF IEEE 15(7) 1029(1963). = + AND J.L. OOLBY,
ET.AL., (ED.), HOLT, RINEHART AND WINSTON, INC., NEW YORK, 1961
ET.AL., (ED.), HOLT, RINEHART AND WINSTON, INC., NEW YORK, 1961
EUROPEAN BIOPHYSICS CONGRESS, 1971. = + A CLASS OF FORMAL LA
EUROPEAN BIOPHYSICS CONGRESS, 1971. = + A CLASS OF FORMAL LA
FACTORS IN CLASSIFICATION," S. FROM J.M. PERREULT, "PROCE
FALL JOINT COMPUTER CONFERENCE 27(11), 325(1963). = ANALYST PR
FALL JOINT COMPUTER CONFERENCE, 27(11), 317-326(1965). = S PR
FALL JOINT COMPUTER CONFERENCE, 27(11), 317-326(1965). = S PR
FALL JOINT COMPUTER CONFERENCE, 307-316(1965). = N ECONOMICA
FALL JOINT COMPUTER CONFERENCE, 307-316(1965). = N ECONOMICA
FALL JOINT COMPUTER CONFERENCE, 33(1), 1968. = R, A COMPUT
FALL JOINT COMPUTER CONFERENCE, 33(1), 1968. = R, A COMPUT
FEDERATION OF INFORMATION PROCESSING CONGRESS, MUNICH, 1962. +
FEDERATION OF INFORMATION PROCESSING CONGRESS, MUNICH, 1962. +
FEDERATION OF INFORMATION PROCESSING CONGRESS, NORTH HOLLAND. +
FEDERATION OF INFORMATION PROCESSING CONGRESS, NORTH HOLLAND. +
FILLMORE, "A PROPOSAL CONCERNING ENGLISH PREPOSITIONS," GEORG
FILLMORE, "IMPROVEMENT IN CASE GRAMMAR, 1970." PRESENTED AT T
FILLMORE, "SOME PROBLEMS FOR CASE GRAMMAR," WORKING PAPERS IN
FILLMORE, "THE CASE FOR CASE," IN UNIVERSALS IN LINGUISTIC TH
FILLMORE, "THE CASE FOR CASE," IN UNIVERSALS IN LINGUISTIC TH
FIRST EUROPEAN BIOPHYSICS CONGRESS, 1971. = + A CLASS OF FOR
FIRST EUROPEAN BIOPHYSICS CONGRESS, 1971. = + A CLASS OF FOR
IDA, APRIL, 1967. = + MEETING OF THE AMERICAN CHEMICAL
O. STOCKHOLM. = + NO TESLER, "CONCEPTUAL DEPENDENCY PARSE"
FORMAL LANGUAGES FOR POSSIBLE BIOLOGICAL APPLICATION AND INFORMA
FORMAL LANGUAGES FOR POSSIBLE BIOLOGICAL APPLICATION, FIRST
FORMALISM AND SOME OBSERVATIONS," LANGUAGE 40(4), 511-525(19
FORMALISM AND SOME OBSERVATIONS," LANGUAGE 40(4), 511-525(19
FOUNDATIONS OF OFFICE OF SCIENCE INFORMATION SERVICE, O.S.U., CO
FOUNDATIONS OF LANGUAGE, 6, 322-361(1970). = AGE AS SEEN FROM
FRANCISCO, CALIFORNIA, 1964. = A. A. KAPLAN, THE C
FRIEDMAN, B.C. WALL AND O.E. WALKER, "THE MITRE SYNAPTIC ANA

- [illegible]

- LANGUAGES FOR POSSIBLE BIOLOGICAL APPLICATION, " FIRST EUROPEA 0036
LANGUAGES FOR POSSIBLE BIOLOGICAL APPLICATION, " FIRST EUROPEA 0014
LANGUAGES, AND SEMIGROUPS, ACADEMIC PRESS, NEW YORK, 1968. P.
LANGUAGES, AND SEMIGROUPS, ACADEMIC PRESS, NEW YORK, 1968. =
LANGUAGES, PERGAMON PRESS, NEW YORK, 1966. = C LINGUISTICS A
LANGUAGES, " IN M.A. ARBIB (ED.), ALGEBRAIC THEORY OF MACHINES 0036
LANGUAGES, " IN M.A. ARBIB (ED.), ALGEBRAIC THEORY OF MACHINES 0036
LANGUAGES, " REPORT TO NATIONAL SCIENCE FOUNDATION OFFICE OF \$ 0014
LAY, THE OUNLE-KWIC COORDINATE INDEX: A NEW APPROACH FOR PRE 0157
LECTURES ON ALGEBRAIC LINGUISTICS AND MACHINETRANSLATION, " IN 0005
LIMITED PARSING OF ENGLISH, " PROC. OF FALL JOINT COMPUTER CO 0056
LIMITED PARSING OF ENGLISH, " PROC. OF FALL JOINT COMPUTER CO 0055
LINGUISTIC SOCIETY OF AMERICA AT THE OHIO STATE UNIVERSITY, " 0137
LINGUISTIC THEORY, BACH, ET-AL. (ED.), HOLY, RINEHART AND W 0014
LINGUISTIC THEORY, BACH, ET-AL. (ED.), HOLY, RINEHART AND W 0134
LINGUISTICS AND MACHINETRANSLATION, " IN AUTOMATIC TRANSLATION 0005
LINGUISTICS WORKING PAPERS NO.1, GEORGETOWN UNIVERSITY PRESS 0137
LINGUISTICS WORKING PAPERS NO.2, GEORGETOWN UNIVERSITY PRESS 0137
LINGUISTICS WORKING PAPERS NO.2, GEORGETOWN UNIVERSITY PRESS 0137
LINGUISTICS WORKING PAPERS NO.6, GEORGETOWN UNIVERSITY PRESS 0137
LINGUISTICS WORKING PAPERS, GEORGETOWN UNIVERSITY PRESS, WAS 0137
LINGUISTICS 3(1), 1967. = M.A.A. HALLIDAY, "NOTES ON TRA 0006
LINGUISTICS, COLING, PREPRINT NO.2, FOCK, STOCKHOLM. = AND TR 0137
LINGUISTICS, HOLY, RINEHART AND WINSTON, NEW YORK, 1961. = 0066
LINGUISTICS, UNIVERSITY PRESS, CARBRIDGE, 1968. = 0004
LINGUISTICS, 3(1), 37-81(1967). = A.K. HALLIDAY, "NOTES ON TRA 0054
LINGUISTICS, 3(2), 200(1967). = A.K. HALLIDAY, "NOTES ON TRA 0054
LINGUISTICS, 6, 31-51(1970) = ER, "A CONCEPTUAL DEPENDENCY P 0164
LINGUISTICS, 19TH MONOGRAPH, 1966. = C ENGLISH PREPOSITIONS, " 0136
LINKS AND ROLES, " AMERICAN DOCUMENTATION 13, 104-11 (1962). = 0137
LITERATURE DIVISION, MIAMI 8FACH, FLORIDA, APRIL, 1967. = AL 0157
LITERATURE, WILFY-INTERSCIENCE, NEW YORK, 1971. = S. SMITH AN 0137
LOCKHEED MISSILES AND SPACE CO., PALO ALTO, CALIFORNIA, 1968. 0070
LOCKHEED MISSILES AND SPACE CO., PALO ALTO, CALIFORNIA, 1967. 0070
LONDON, 1964, P.219. = J. HONELL, DISCOURSE CONCERNING THE 0156
LONDON, 1961, IN C. FRIES, THE STRUCTURE OF ENGLISH, HARCOURT 0065
LONDON, 1974 IN C. FRIES, THE STRUCTURE OF ENGLISH, HARCOURT 0065
LOOKING GLASS, AIRMONT PUBLISHING CO., NEW YORK, 1965, P.141. 0036
LOOKING GLASS, " FROM ALICE'S ADVENTURES IN WONDERLAND AND THR 0066
LOVE, "NOTICE-ACCESS MEMORY RETRIEVAL USING TRUNCATED RECORD 0049
LYNCH, "ARTICULATION IN THE GENERATION OF SUBJECT INDEXES BY 0157
LYONS, INTRODUCTION TO THEORETICAL LINGUISTICS, UNIVERSITY PR 0006
M.A. ARBIB (ED.), ALGEBRAIC THEORY OF MACHINES, LANGUAGES, A 0036
M.A. ARBIB (ED.), ALGEBRAIC THEORY OF MACHINES, LANGUAGES, A 0001
M.A.K. HALLIDAY, "FUNCTIONAL DIVERSITY IN LANGUAGE AS SEEN FR 0054
M.A.K. HALLIDAY, "NOTES ON TRANSITIVITY AND THEME IN ENGLISH, 0054
M.A.K. HALLIDAY, "NOTES ON TRANSITIVITY AND THEME IN ENGLISH, 0054
M.A.K. HALLIDAY, "NOTES ON TRANSITIVITY AND THEME IN ENGLISH, 0054
M.F. LYNCH, "ARTICULATION IN THE GENERATION OF SUBJECT INDEXE 0157
M.R. QUILLIAN, "SEMANTIC MEMORY, " IN MI MINSKY (ED.) SEMANTIC 0061
MACHINE INTELLIGENCE 3, D. HECHIE (ED.), P.201-309. = OR, 0062
MACHINE INTELLIGENCE 3, D. HECHIE (ED.), P.201-309. = OR, 0062
MACHINE TRANSLATION OF LANGUAGES AND APPLIED LANGUAGE ANALY 0166
MACHINE, " IN MACHINE INTELLIGENCE 3, D. HECHIE (ED.), P.307 0062
MACHINE, " IN MACHINE INTELLIGENCE 3, D. HECHIE (ED.), PP281- 0061
MACHINES, LANGUAGES, AND SEMIGROUPS, ACADEMIC PRESS, NEW YOR 0006
MACHINES, LANGUAGES, AND SEMIGROUPS, ACADEMIC PRESS, NEW YOR 0006

- ◀LEL, "FOUR LECTURES ON ALGEBRAIC LINGUISTICS AND
IND. "SEMANTIC ANALYSIS PROCEDURES IN THE SCIENCES OF
XT." PASTERS THESIS FROM THE OHIO STATE UNIVERSITY,
C INFORMATION PROCESSING, THE MIT PRESS, CAMBRIDGE,
EGTS OF THE THEORY OF SYNTAX. MIT PRESS, CAMBRIDGE,
DESCRIPTIVE INTRODUCTION, THE MIT PRESS, CAMBRIDGE,
TER PROGRAMMING, ADDISON-WESLEY PUBLISHING COMPANY,
GENERATING STRUCTURAL SURROGATES OF ENGLISH TEXT.
1964. PP.230. = Z. GRAMMAR
ETOWN UNIVERSITY + W.A. COOK, S.J., "A CASE GRAMMAR
AGO PRESS, 1970. =
W.L. CHAFFE,
ENGLISH BY MACHINE," IN MACHINE INTELLIGENCE 3. O.
ENGLISH BY MACHINE," IN MACHINE INTELLIGENCE 3. O.
UAL NATIONAL COLLOQUIUM ON INFORMATION RETRIEVAL,
EXES BY COMPUTER. PRESENTED AT THE 193RD NATIONAL
SEPTMBER, 28-31(1967). + T.L. LONE, "DIRECT ACCESS
THE MIT PRESS, CAMBRIDGE. + M.R. QULLIAN, "SEMANTIC
ES IN RELATIONAL ASPECTS OF CONCEPTUALIZATION IN
OF CONCEPTS," ANGEWANDTE CHEMIE, 9(+ TOPOLOGICAL
+ NICKELSEN AND J.H. WINTER, "TOSAR - A TOPOLOGICAL
+ NICKELSEN AND J.H. WINTER, "TOSAR - A TOPOLOGICAL
+ NICKELSEN AND J.H. WINTER, "TOSAR - A TOPOLOGICAL
+ NICKELSEN AND J.H. WINTER, "TOSAR - A TOPOLOGICAL
DEPENDENCY PARSE FOR NATURAL LANGUAGE," A STATISTICAL
CAH CHEMICAL SOCIETY, CHEMICAL LITERATURE DIVISION,
CAMBRIDGE. + M.P. QULLIAN, "SEMANTIC MEMORY," IN MI
NISKY (ED.), SEMANTIC INFORMATION PROCESSING, THE MIT
ORT: AUTOMATIC INDEXING AND ABSTRACTING. LOCKHEED
CKY. J. FRIEDMAN, B.C. HALL AND D.E. WALKER, "THE
CKY. J. FRIEDMAN, B.C. HALL AND D.E. WALKER, "THE
CKY. J. FRIEDMAN, B.C. HALL AND D.E. WALKER, "THE
RSITY IN LANGUAGE AS SEEN FROM A CONSIDERATION OF
WORK, "A THEORY OF INDEXING: INDEXING THEORY AS A
NORY, "A THEORY OF INDEXING: INDEXING THEORY AS A
MONMS. J. BURGER AND R. SCHWARZ, "A COMPUTATIONAL
MONMS. J. BURGER AND R. SCHWARZ, "A COMPUTATIONAL
RSTY ROUND TABLE ON LANGUAGES AND LINGUISTICS, 19TH
S AND ROLES." AMERICAN DOCUMENTATION 13, 104-+ B.A.
OCIETY FOR INFORMATION SCIENCE 23(3), 195-219+ C.A.
+UAGE AS SEFFI FROM A CONSIDERATION OF MODALITY AND
NALAL OF CHEMICAL DOCUMENTATION 10+ H. SKOLNIK, "THE
RIVAL," JOURNAL OF CHEMICAL DOCU+ H. SKOLNIK, "THE
ONAL FEDERATION OF INFORMATION PROCESSING CONGRESS,
PROCESSING CONGRESS, NORTH HOLLAND PUBLISHING CO.
PROCESSING CONGRESS, NORTH HOLLAND PUBLISHING CO.
TION, THE MIT PRESS, CAMBRIDGE. MASSACHUSETTS, 1964
RECT-ACCESS MEMORY RETRIEVAL USING TRUNCATED RECORD
UNCTION WORDS." FROM A.O. BERTON ED.) 7TH ANNUAL
BJECT INDEXES BY COMPUTER. PRESENTED AT THE 193RD
VIS. J. ROTHSTEIN, "RELATIONAL LANGUAGES," REPORT TO
V. 591+ W.A. WOODS, "TRANSITION NETWORK GRAMMARS FOR
V. 591+ W.A. WOODS, "TRANSITION NETWORK GRAMMARS FOR
V. 591+ W.A. WOODS, "TRANSITION NETWORK GRAMMARS FOR
V. 591+ W.A. WOODS, "TRANSITION NETWORK GRAMMARS FOR
- MACHINETranslation." In Automatic Translation of Languages. + 0009
MAITZELL, J.F. SHYTH AND T.E. SINGER, ABSTRACTING SCIENTIFIC A 0137
MAN." SOCIAL SCIENCE INFORMATION, 8(13), 17-45(1969). = + GARD 0013
MARCH, 1973. = +GENERATING STRUCTURAL SURROGATES OF ENGLISH TE 0036
MASS., 1968. = +SEMANTIC MEMORY," IN MI MINSKY (EO.) SEMANTI 0137
MASS. 1969. =
MASSACHUSETTS, 1968. = + CATTELL, THE NEW ENGLISH GRAMMAR: A 0006
MASSACHUSETTS, 1968. = O.E. KNUTH, THE ART OF COMPU 0001
MASTERS THESIS FROM THE OHIO STATE UNIVERSITY, MARCH. 1973. = 0036
MATHEMATICAL STRUCTURE OF LANGUAGE, INTERSCIENCE PUBLISHERS, 0033
MATRIX," LANGUAGES AND LINGUISTICS WORKING PAPERS NO.6, GEORG 0137
MEANING AND THE STRUCTURE OF LANGUAGE, THE UNIVERSITY OF CHIC 0137
MECHIE (ED.), P.307 AND 308. = +R, "THE SYNTACTIC ANALYSIS OF 0062
MEDIC (EO.), PP281-309. = +DEWAR, "THE SYNTACTIC ANALYSIS OF 0063
MEETING OF THE AMERICAN CHEMICAL SOCIETY, PHILADELPHIA, PENNSYLVANIA, 1+ 0095
MEMORY RETRIEVAL USING TRUNCATED RECORD NAMES," SOFTWARE AGE, 0066
MEMORY," IN MI MINSKY (EO.) SEMANTIC INFORMATION PROCESSING. 0157
MESSAGA ANALYSIS. FROM J.M. PERREAU, "PROCEEDINGS OF THE I+ 0169
METHOD FOR REPRESENTATION OF SYNTHETIC AND ANALYTICAL RELATI 0169
METHOD FOR REPRESENTATION OF SYNTHETIC AND ANALYTICAL RELATI 0161
METHOD FOR REPRESENTATION OF SYNTHETIC AND ANALYTICAL RELATI 0156
METHOD FOR REPRESENTATION OF SYNTHETIC AND ANALYTICAL RELATI 0157
METHOD FOR REPRESENTATION OF SYNTHETIC AND ANALYTICAL RELATI 0164
METHODS IN LINGUISTICS, 6, 31-51(1970) = +ER, "A CONCEPTUAL 0164
MIAMI BEACH, FLORIDA, APRIL, 1967. = +AL MEETING OF THE AMERI 0157
MINSKY (EO.) SEMANTIC INFORMATION PROCESSING, THE MIT PRESS, 0157
MISSILES AND SPACE CO., PALO ALTO, CALIFORNIA, 1967 (AO 647 + 0070
MISSILES AND SPACE CO., PALO ALTO, CALIFORNIA, 1967 (AO 659 + 0070
MITRE SYNTACTIC ANALYSIS PROCEDURE FOR TRANSFORMATIONAL GRAM+ 0112
MITRE SYNTACTIC ANALYSIS PROCEDURE FOR TRANSFORMATIONAL GRAM+ 0052
MITRE SYNTACTIC ANALYSIS PROCEDURE FOR TRANSFORMATIONAL GRAM+ 0053
MODALITY AND MOOD IN ENGLISH," FOUNDATIONS OF LANGUAGE, 6, 3+ 0054
MODEL FOR INFORMATION STORAGE AND RETRIEVAL," COMPUTER AND IN+ 0158
MODEL FOR INFORMATION STORAGE AND RETRIEVAL," COMPUTER AND IN+ 0160
MODEL OF VERBAL UNDERSTANDING," PROC OF THE FALL JOINT COMPU+ 0012
MODEL OF VERBAL UNDERSTANDING," PROC OF THE FALL JOINT COMPU+ 0014
MOONOGRAPH, 1964. = +G ENGLISH PREPOSITIONS," GEORGETOWN UNIV+ 0136
MONTAGUE, "PATENT INDEXING BY CONCEPT COORDINATION USING LINK 0137
MONTAGUE, "LINGUISTICS AND INFORMATION SCIENCE," AMERICAN 0014
MOOD IN ENGLISH," FOUNDATIONS OF LANGUAGE, 6, 322-361(1970).+ 0054
MULTITERM INDEX: A NEW CONCEPT IN INFORMATION RETRIEVAL," JOU 0137
MULTITERM INDEX: A NEW CONCEPT IN INFORMATION STORAGE AND RET 0157
MUNICH, 1962. = +SYNTACTIC ANALYZER," PROC. OF THE INTERNATI 0112
MUNICH, 1962. = +H SYNTACTIC ANALYZER," PROC OF THE INTERNATI 0066
MUNICH, 1962, 306-312. = +RNATIONAL FEDERATION OF INFORMATION 0051
MUNICH, 1962, 306-312. = +RNATIONAL FEDERATION OF INFORMATION 0051
MUNICH, 1962, 306-312. = +RNATIONAL FEDERATION OF INFORMATION 0051
N.R. CATTELL, THE NEW ENGLISH GRAMMAR: A DESCRIPTIVE INTRODU 0005
NAMES." SOFTWARE AGE, SEPTEMBER, 28-31(1967). = +L. LONE, "DI 0066
NATIONAL COLLOQUIUM ON INFORMATION RETRIEVAL, MEDICAL OCCUPA+ 0095
NATIONAL MEETING OF THE AMERICAN CHEMICAL SOCIETY, CHEMICAL + 0157
NATIONAL SCIENCE FOUNDATION OFFICE OF SCIENCE INFORMATION SER 0014
NATURAL LANGUAGE ANALYSIS," COMMUNICATIONS OF THE ACM, 13(10) 0112
NATURAL LANGUAGE ANALYSIS," COMMUNICATIONS OF THE ACM, 13(10) 0112
NATURAL LANGUAGE ANALYSIS," COMMUNICATIONS OF THE ACM, 13(10) 0064
NATURAL LANGUAGE ANALYSIS," COMMUNICATIONS OF THE ACM, 13(10) 0137

- T. WINOGRAD, "UNDERSTANDING
T. WINOGRAD, "UNDERSTANDING
T. WINOGRAD, "UNDERSTANDING
T. WINOGRAD, "UNDERSTANDING
AND L. TESLER, "A CONCEPTUAL DEPENDENCY PARSER FOR
M. SAGER AND M. SALKOFF, "SYNTACTIC ANALYSIS OF
N. SAGER AND M. SALKOFF, "SYNTACTIC ANALYSIS OF
NS OF THE ACM, 13(10), 591+ M.A. WOODS, "TRANSITION
NS OF THE ACM, 13(10), 591+ M.A. WOODS, "TRANSITION
NS OF THE ACM, 13(10), 591+ M.A. WOODS, "TRANSITION
NS OF THE ACM, 13(10), 591+ M.A. WOODS, "TRANSITION
REPRESENTATION OF SYNT. R. FUGHMAN, H. NICKELSEN, I.
REPRESENTATION OF SYNT. R. FUGHMAN, H. NICKELSEN, I.
REPRESENTATION OF SYNT. R. FUGHMAN, H. NICKELSEN, I.
L METHOD FOR REPRESENTATION OF SYNT. R. FUGHMAN, H.
AL METHOD FOR REPRESENTATION OF SYNT. R. FUGHMAN, H.
AL METHOD FOR REPRESENTATION OF SYNT. R. FUGHMAN, H.
RULES," LANGUAGES AND LINGUISTICS WORKING PAPERS
MS FOR CASE GRAMMAR," WORKING PAPERS IN LINGUISTICS,
"ALYSIS," LANGUAGES AND LINGUISTICS WORKING PAPERS
"MATRIX," LANGUAGES AND LINGUISTICS WORKING PAPERS
ONAL FEDERATION OF INFORMATION PROCESSING CONGRESS,
ONAL FEDERATION OF INFORMATION PROCESSING CONGRESS,
ONAL FEDERATION OF INFORMATION PROCESSING CONGRESS,
E FOUNDATION OFFICE OF SCIENCE INFORMATION SERVICE,
G. HAYES, "DEPENDENCY THEORY: A FORMALISM AND SOME
G. HAYES, "DEPENDENCY THEORY: A FORMALISM AND SOME
G. HAYES, "DEPENDENCY THEORY: A FORMALISM AND SOME
OC. OF THE INTERNATIONAL FEDERATION OF INFORMAT*, A.
OC. OF THE INTERNATIONAL FEDERATION OF INFORMAT*, A.
OC. OF THE INTERNATIONAL FEDERATION OF INFORMAT*, A.
OC. OF THE INTERNATIONAL FEDERATION OF INFORMAT*, A.
"LANGUAGES," REPORT TO NATIONAL SCIENCE FOUNDATION
PROGRESSES OF ENGLISH TEXT," MASTERS THESIS FROM THE
MAJON SCIENCE RESEARCH CENTER, CISCRC-TR-71-13, THE
STITUTE OF THE LINGUISTIC SOCIETY OF AMERICA AT
GRAMMAR," WORKING PAPERS IN LINGUISTICS, NT-10, THE
MATION SCIENCE RESEARCH CENTER, CISCRC-TR-71-13, THE
H-W. BUTTELMANN, PRIVATE COMMUNICATION, THE
E OF SCIENCE INFORMATION SERVICE, O.S.U., COLUMBUS,
BERKELEY, 1962." S.M. LAMB,
WORKING CLASS, ARMY PUBLISHING CO., NEW YORK, 1964,
KINGS, PRINTED FOR RONLAND REVOLDS, LONDON, 1964,
CHINE," IN MACHINE INTELLIGENCE 3, D. MECHE (ED.),
CAGNE," HOLY, RINEHART AND WINSTON, NEW YORK, 1933,
R, CISCRC-TR-71-13, THE OHIO STATE UNIVERSITY, 1911,
R, CISCRC-TR-71-13, THE OHIO STATE UNIVERSITY, 1911,
R, CISCRC-TR-71-13, THE OHIO STATE UNIVERSITY, 1911,
HOLT, RINEHART AND WINSTON, INC., NEW YORK, 1969,
HOLT, RINEHART, AND WINSTON, INC., NEW YORK, 1969,
ISH, HARLCOURT, BRACE & WORLD, INC., NEW YORK, 1952,
LISH, HARLCOURT, BRACE & WORLD, INC., NEW YORK, 1952,
ISH, HARLCOURT, BRACE & WORLD, INC., NEW YORK, 1952,
SHAXE] STRUCTURAL, PARIS, KLENHSUCK, QUOTED BY R.
- NATURAL LANGUAGE," COGNITIVE PSYCHOLOGY, 3, 1-191 (1972). = 0137
NATURAL LANGUAGE," COGNITIVE PSYCHOLOGY, 3, 1-191(1972). = 0112
NATURAL LANGUAGE," COGNITIVE PSYCHOLOGY, 3, 1-191(1972). = 0058
NATURAL LANGUAGE," COGNITIVE PSYCHOLOGY, 3, 1-191(1972). = 0055
NATURAL LANGUAGE," STATISTICAL METHODS IN LINGUISTICS, 6, 31+ 0006
NATURAL LANGUAGE," ADVANCES IN COMPUTERS R, 153-188(1969). = 0164
NATURAL LANGUAGE," ADVANCES IN COMPUTERS R, 153-188(1969). = 0054
NETWORK GRAMMARS FOR NATURAL LANGUAGE ANALYSIS," COMMUNICATIO 0137
NETWORK GRAMMARS FOR NATURAL LANGUAGE ANALYSIS," COMMUNICATIO 0112
NETWORK GRAMMARS FOR NATURAL LANGUAGE ANALYSIS," COMMUNICATIO 0112
NICATIONS OF THE ACM, 5(2), 103-144(1962). = 0043
NICKELSEN AND J.H. WINTER, "TOSAR - A TOPOLOGICAL METHOD FOR 0157
NICKELSEN AND J.H. WINTER, "TOSAR - A TOPOLOGICAL METHOD FOR 0157
NICKELSEN AND J.H. WINTER, "TOSAR - A TOPOLOGICAL METHOD FOR 0157
NICKELSEN, I. NICKELSEN AND J.H. WINTER, "TOSAR - ATOLOGICA 0161
NICKELSEN, I. NICKELSEN AND J.H. WINTER, "TOSAR - ATOLOGICA 0161
NICKELSEN, I. NICKELSEN AND J.H. WINTER, "TOSAR - A TOPOLOGIC 0164
NICKELSEN, I. NICKELSEN AND J.H. WINTER, "TOSAR - A TOPOLOGIC 0157
NICKELSEN, I. NICKELSEN AND J.H. WINTER, "TOSAR - A TOPOLOGIC 0156
NO.10, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, D.C., 1970, P. 0137
NO.2, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, D.C., 1971, P. 0137
NO.6, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, D.C., 1972, P. 0137
NORTH HOLLAND PUBLISHING CO., MUNICH, 1962, 306-312. = +RNATI 0051
NORTH HOLLAND PUBLISHING CO., MUNICH, 1962, 306-312. = +RNATI 0051
O.S.U., COLUMBUS, OHIO, 1970. = "REPORT TO NATIONAL SCIENCE 0014
OBSERVATIONS," LANGUAGE, 40(4), 511-525(1964). = 0 0006
OBSERVATIONS," LANGUAGE, 40(4), 511-525(1964). = 0 0162
OETTINGER AND J. KUND, "MULTIPLE-PATH SYNTACTIC ANALYZER," PR 0051
OETTINGER AND S. KUND, "MULTIPLE-PATH SYNTACTIC ANALYZER," PR 0051
OETTINGER AND S. KUND, "MULTIPLE-PATH SYNTACTIC ANALYZER," PR 0051
OFFICE OF SCIENCE INFORMATION SERVICE, O.S.U., COLUMBUS, OHI 0014
OHIO STATE UNIVERSITY, MARCH, 1973. = +GENERATING STRUCTURAL S 0036
OHIO STATE UNIVERSITY, 1911, P.53. = +VAL COMPUTER AND INFOR 0160
OHIO STATE UNIVERSITY, 1970. = +, PRESENTED AT THE SUMMER IN 0137
OHIO STATE UNIVERSITY, 1971. P.53. = +SOME PROBLEMS FOR CASE 0137
OHIO STATE UNIVERSITY, 1973. = 0112
OHIO, 1970. = +, REPORT TO NATIONAL SCIENCE FOUNDATION OFFIC 0014
OUTLINE OF STRATIFICATION GRAMMAR, UNIVRSITY OF CALIFORNIA, 0006
P.219. = +J'S ADVENTURES IN WONDERLAND AND THROUGH THE LOO 0069
P.307 AND 308. = +R, "THE SYNTACTIC ANALYSIS OF ENGLISH BY MA 0062
P.5. = +, BLOOMFIELD, LA 0065
P.53. = +AL," COMPUTER AND INFORMATION SCIENCE RESEARCH CENTE 0160
P.65. = +VAL," COMPUTER AND INFORMATION SCIENCE RESEARCH CENTE 0158
P.65. = +W.A. COOK, S.J., INTRODUCTION TO TAGMEMIC ANALYSIS 0020
P.70. = +W.A. COOK, S.J., INTRODUCTION TO TAGMEMIC ANALYSIS, 0036
P.74. = +C. PRIES, THE STRUCTURE OF ENGL 0069
P.75. = +C. PRIES, THE STRUCTURE OF ENGL 0066
PAGES IN "RELATIONAL ASPECTS OF CONCEPTUALIZATION IN MESSAGE 0169

- G AND ABSTRACTING, LOCKHEED MISSILES AND SPACE CO.,
INDEXING, LOCKHEED MISSILES AND SPACE CO.,
FILLMORE, "SOME PROBLEMS FOR CASE GRAMMAR," WORKING
PAPERS NO. 1, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 2, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 3, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 4, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 5, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 6, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 7, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 8, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 9, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 10, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 11, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 12, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 13, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 14, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 15, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 16, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 17, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 18, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 19, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 20, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 21, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 22, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 23, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 24, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 25, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 26, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 27, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 28, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 29, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 30, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 31, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 32, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 33, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 34, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 35, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 36, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 37, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 38, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 39, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 40, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 41, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 42, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 43, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 44, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 45, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 46, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 47, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 48, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 49, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 50, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 51, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 52, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 53, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 54, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 55, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 56, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 57, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 58, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 59, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 60, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 61, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 62, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 63, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 64, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 65, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 66, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 67, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 68, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 69, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 70, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 71, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 72, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 73, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 74, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 75, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 76, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 77, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 78, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 79, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 80, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 81, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 82, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 83, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 84, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 85, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 86, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 87, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 88, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 89, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 90, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 91, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 92, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 93, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 94, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 95, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 96, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 97, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 98, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 99, LANGUAGES AND LINGUISTICS WORKING
PAPERS NO. 100, LANGUAGES AND LINGUISTICS WORKING
- PALO ALTO, CALIFORNIA, 1967 (AD 659 057). = AUTOMATIC INDEXIN 0070
PALO ALTO, CALIFORNIA, 1968 (AD 667 473). = RHATIVE ABSTRACT 0070
PAPERS IN LINGUISTICS, NO.10, THE OHIO STATE UNIVERSITY, 1973, 0137
PAPERS NO.1, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, O.C., 0137
PAPERS NO.2, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, O.C., 0137
PAPERS NO.3, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, O.C., 0137
PAPERS NO.4, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, O.C., 0137
PAPERS NO.5, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, O.C., 0137
PAPERS NO.6, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, O.C., 0137
PAPERS NO.7, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, O.C., 0137
PAPERS NO.8, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, O.C., 0137
PAPERS NO.9, GEORGETOWN UNIVERSITY PRESS, WASHINGTON, O.C., 0137
PARIS, KLENCHSKY, QUOTED BY R. PAGES IN "RELATIONAL ASPECTS 0166
PARSER FOR NATURAL LANGUAGE," STATISTICAL METHODS IN LINGUIS 0166
PARSER, COMPUTATIONAL LINGUISTICS, COLING, PREPRINT NO. 2, F 0137
PARSING OF ENGLISH," PROC. OF FALL JOINT COMPUTER CONFERENCE 0056
PARSING OF ENGLISH," PROC. OF FALL JOINT COMPUTER CONFERENCE 0055
PART 1," JOURNAL OF LINGUISTICS, 3(1), 37-81(1967). = K. MAL 0054
PART 1," JOURNAL OF LINGUISTICS, 3(1), 1967, M.A.K. MAL 0054
PART 2," JOURNAL OF LINGUISTICS, 3(2), 200(1967). = A.K. MAL 0054
PARTS OF SPEECH IN ENGLISH," PROC. OF IEEE 19(17), 1029(1963) 0062
PARTS OF SPEECH IN ENGLISH," PROC. OF IEEE 19(17), 1029(1963) 0062
PENNSYLVANIA, 1970, PP.71-81. = LLOQUIM ON INFORMATION RETR 0005
PERGAMON PRESS, NEW YORK, 1966. = C LINGUISTICS AND MACHINE 0005
PERRAULT, "PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON RE 0189
PETRARCA AND W.M. LAY, THE ORUBLE-KNIC COORDINATE INDEX: A NE 0189
PHILADELPHIA, PENNSYLVANIA, 1970, PP.71-81. = LLOQUIM ON I 0166
PLATH, "AUTOMATIC SENTENCE DIAGRAMMING," 1961 INTERNATIONAL 0166
POSSIBLE BIOLOGICAL APPLICATION," FIRST EUROPEAN BIOPHYSICS C 0036
PP.1-9. = L LANGUAGES AND LINGUISTICS WORKING PAPERS NO.2, G 0137
PP.10-23. = L LANGUAGES AND LINGUISTICS WORKING PAPERS NO.2, G 0137
PP.14-30. = L LANGUAGES AND LINGUISTICS WORKING PAPERS NO.2, G 0137
PP.15-48. = L LANGUAGES AND LINGUISTICS WORKING PAPERS NO.2, G 0137
PP.230. = Z. HARRIS, MATHEMATICAL ST 0053
PP.313-327. = B (EO.), ALGEBRAIC THEORY OF MACHINES, LANGUA 0036
PP.71-81. = LLOQUIM ON INFORMATION RETRIEVAL, MEDICAL OCUM 0036
PREDECIENCY OF KINGS, PRINTED FOR ROWLAND REYNOLDS, LONDON, 16 0061
PREPARATION OF HIGH-QUALITY INDEXES BY AUTOMATIC INDEXING TE 0157
PREPOSITIONS," GEORGETOWN UNIVERSITY ROUND TABLE ON LANGUAGES 0136
PREPRINT FOR LANGUAGES AND LINGUISTICS WORKING PAPERS, GEORGE 0136
PREPRINT NO. 2, FOCK, STOCKHOLM. = NO TESLER, "CONCEPTUAL DE 0137
PRESENT-DYAMERICAN ENGLISH, BROOKLYN UNIVERSITY, PROVIDENCE, R 0068
PRESENTED AT SHARE XXVII, NEW YORK CITY, AUGUST 12, 1971. = 0037
PRESENTED AT THE SUMMER INSTITUTE OF THE LINGUISTIC SOCIETY 0137
PRESS, CAMBRIDGE, MASS., 1968. = SENSEMANTIC MEMORY," IN MI 0157
PRESS, CAMBRIDGE, MASS., 1965.
PRESS, CAMBRIDGE, MASSACHUSETTS, 1966. = CATTELL, THE NEW E 0005
PRESS, CAMBRIDGE, MASSACHUSETTS, 1966. = CATTELL, THE NEW E 0005
PRESS, CAMBRIDGE, MASSACHUSETTS, 1966. = CATTELL, THE NEW E 0005
PRESS, COLORADO, 1972. = P. RECKMAN, T 0066
PRESS, NEW YORK, 1966. = N M.A. ARRIO (EO.), ALGEBRAIC THEO 0005
PRESS, NEW YORK, 1968. = N M.A. ARRIO (EO.), ALGEBRAIC THEO 0036
PRESS, NEW YORK, 1968. PP.313-327. = B (EO.), PREPRINT FOR LANGUAGES AN 0137
PRESS, WASHINGTON, O.C., 1970, PP.14-30. = L LANGUAGES AND LIN 0137
PRESS, WASHINGTON, O.C., 1971, PP.1-9. = L LANGUAGES AND LIN 0137
PRESS, WASHINGTON, D.C., 1971, PP.10-23. = L LANGUAGES AND LIN 0137
PRESS, WASHINGTON, D.C., 1972, PP.15-48. = L LANGUAGES AND LIN 0137
PRESS, WASHINGTON, D.C., 1970. = W.L. CHAFE, MEANING AND T 0012
PRESS, 1970. =

PRIESTLY, RUDIMENTS OF ENGLISH GRAMMAR, LONDON, 1761, IN C. F. 0085
PRIVATE FOR ROMLAND REYNOLDS, LONDON, 1664, P.219. = J. MO 0112
PROBLEMS FOR CASE GRAMMAR." WORKING PAPERS IN LINGUISTICS, NO. 0137
PROC OF THE FALL JOINT COMPUTER CONFERENCE, 33(1), 1968. = *R 0014
PROC OF THE FALL JOINT COMPUTER CONFERENCE, 33(1), 1968. = *R 0012
PROC OF THE INTERNATIONAL FEDERATION OF INFORMATION PROCESS+ 0006
PROC OF THE INTERNATIONAL FEDERATION OF INFORMATION PROCESS+ 0055
PROC. OF FALL JOINT COMPUTER CONFERENCE, 307-316(1965). = *N 0055
PROC. OF FALL JOINT COMPUTER CONFERENCE, 307-316(1965). = *N 0056
PROC. OF IEEE 15(7) 1029(1963). = *AND J.L. DOLBY, "AUTOMATI 0063
PROC. OF IEEE 15(7), 1029(1963). = *AND J.L. DOLBY, "AUTOMATI 0062
PROC. OF THE FALL JOINT COMPUTER CONFERENCE, 27(1), 325(1965)+ 0053
PROC. OF THE FALL JOINT COMPUTER CONFERENCE, 27(1), 317-326(+ 0052
PROC. OF THE FALL JOINT COMPUTER CONFERENCE, 27(1), 317-326(+ 0112
PROC. OF THE INTERNATIONAL FEDERATION OF INFORMATION PROCESS+ 0051
PROC. OF THE INTERNATIONAL FEDERATION OF INFORMATION PROCESS+ 0051
PROC. OF THE INTERNATIONAL FEDERATION OF INFORMATION PROCESS+ 0051
PROCEDURE FOR TRANSFORMATIONAL GRAMMARS, PROC. OF THE FALL J+ 0053
PROCEDURE FOR TRANSFORMATIONAL GRAMMARS, PROC. OF THE FALL J+ 0052
PROCEDURE FOR TRANSFORMATIONAL GRAMMARS," PROC. OF THE FALL + 0112
PROCEEDINGS IN THE SCIENCES OF MAN," SOCIAL SCIENCE INFORMATION 0159
PROCEEDINGS, 24, 365-387(1963). = *ROW, "SYNTACTIC ANALYSIS O 0059
PROCESSING CONGRESS, MUNICH, 1962. = *H SYNTACTIC ANALYZER," P 0112
PROCESSING CONGRESS, MUNICH, 1962. = *H SYNTACTIC ANALYZER," 0006
PROCESSING CONGRESS, NORTH HOLLAND PUBLISHING CO., MUNICH, 1+ 0051
PROCESSING CONGRESS, NORTH HOLLAND PUBLISHING CO., MUNICH, 1+ 0051
PROCESSING CONGRESS, NORTH HOLLAND PUBLISHING CO., MUNICH, 1+ 0051
PROGRAM AND THEORY OF LANGUAGE," COMMUNICATION OF THE ACM, 1+ 0137
PROGRAM FOR LIMITED PARSENG OF ENGLISH," PROC. OF FALL JOINT 0055
PROGRAM FOR LIMITED PARSENG OF ENGLISH," PROC. OF FALL JOINT 0056
PROGRAMMING, ADISON-WESLEY PUBLISHING COMPANY, MASSACHUSETTS 0011
PROPOSAL CONCERNING ENGLISH PREPOSITIONS," GEORGETOWN UNIVERS 0136
PROVIDENCE, R.I., 1967. = *R.R. FRANCIS, COMPUTATIONAL ANALYSI 0068
PSYCHOLOGY, 3, 1-191 (1972). = T. WI 0137
PSYCHOLOGY, 3, 1-191(1972). = T. WI 0112
PSYCHOLOGY, 3, 1-191(1972). = T. WI 0054
PSYCHOLOGY, 3, 1970. = T. WI 0006
PSYCHOLOGY, 3, 191(1972). = T. WI 0055
PUBLISHERS, 1968, PP.230. = Z. HARRI 0053
PUBLISHING CO., MUNICH, 1962, 306-312. = *RNATIONAL FEDERATIO 0051
PUBLISHING CO., MUNICH, 1962, 306-312. = *RNATIONAL FEDERATIO 0051
PUBLISHING CO., MUNICH, 1962, 306-312. = *RNATIONAL FEDERATIO 0051
PUBLISHING CO., NEW YORK, 1965, P.141. = *LICE'S ADVENTURES I 0069
PUBLISHING COMPANY, MASSACHUSETTS, 1968. = O.E. KNU 0011
PUBLISHING COMPANY, SAN FRANCISCO, CALIFORNIA, 1964. = 0010
QUILLIAN, "SEMANTIC MEMORY," IN MI MINSKY (ED.) SEMANTIC INFO 0150
QUILLIAN, "THE TEACHABLE LANGUAGE COMPREHENDER: A SIMULATION 0137
QUOTED BY R. PAGES IN "RELATIONAL ASPECTS OF CONCEPTUALIZAT+ 0169
R.C. SHANK AND L. TESLER, "A CONCEPTUAL DEPENDENCY PARSER FO 0164
R.E. GRISHALO, SNOBOL4 - STRUCTURE AND IMPLEMENTATION, PRESEN 0037
R.E. MAIZELL, J.F. SMITH AND T.F. SINGER, ABSTRACTING SCIENTI 0137
R.E. WALL, "AN ECONOMIC PROGRAM FOR LIMITED PARSENG OF ENGL 0056
R.E. WALL, "AN ECONOMIC PROGRAM FOR LIMITED PARSENG OF ENGL 0055
R.I., 1967. = *R. FRANCIS, COMPUTATIONAL ANALYSIS OF PRESENT 0068
RECORD NAMES," SOFTWARE AGE, SEPTEMBER, 28-33(1967). = *L.O 0066
RELATION OF CONCEPTS," ANGEWANDT CHEMIE, 9(18)59-595(1970)+ 0165

IO. THE OHIO STATE UNIVERSITY. C.-J. FILLMORE, "SOME
H.M. BUTTMANN, "SOME
Z., "A COMPUTATIONAL MODEL OF VERBAL UNDERSTANDING."
AND S. KUNO, "MULTIPLE-PATH SYNTACTIC ANALYZER."
ECONOMICAL PROGRAM FOR LIMITED PARSENG OF ENGLISH."
ECONOMICAL PROGRAM FOR LIMITED PARSENG OF ENGLISH."
C DETERMINATION OF THE PARTS OF SPEECH IN ENGLISH."
C DETERMINATION OF THE PARTS OF SPEECH IN ENGLISH."
ANALYSIS PROCEDURE FOR TRANSFORMATIONAL GRAMMARS.
ANALYSIS PROCEDURE FOR TRANSFORMATIONAL GRAMMARS.
ANALYSIS PROCEDURE FOR TRANSFORMATIONAL GRAMMARS."
AND S. KUNO, "MULTIPLE-PATH SYNTACTIC ANALYZER."
AND S. KUNO, "MULTIPLE-PATH SYNTACTIC ANALYZER."
AND S. KUNO, "MULTIPLE-PATH SYNTACTIC ANALYZER."
ALL AND O.E. WALKER, "THE HITTE SYNTACTIC ANALYSIS
ALL AND O.E. WALKER, "THE HITTE SYNTACTIC ANALYSIS
ALL AND O.E. WALKER, "THE HITTE SYNTACTIC ANALYSIS
F. ENGLISH BY COMPUTER. = J. GARDIN, "SEMANTIC ANALYSIS
OF THE INTERNATIONAL FEDERATION OF INFORMATION
PROC OF THE INTERNATIONAL FEDERATION OF INFORMATION
PROC OF THE INTERNATIONAL FEDERATION OF INFORMATION
C-C OF THE INTERNATIONAL FEDERATION OF INFORMATION
C-C OF THE INTERNATIONAL FEDERATION OF INFORMATION
C-C OF THE INTERNATIONAL FEDERATION OF INFORMATION
IC MEMORY." IN MI MINSKY (ED.) SEMANTIC INFORMATION
COMPUTER+ O.C. CLARKE AND R.E. WALL, "AN ECONOMIC
COMPUTER+ O.C. CLARKE AND R.E. WALL, "AN ECONOMIC
1968. = O.E. KNUTH, THE ART OF COMPUTER
ITY ROUND TABLE ON LANGUAGES AND+ C.J. FILLMORE, "A
OF PRESENT-DAY AMERICAN ENGLISH, BROWN UNIVERSITY,
NOPAD, "UNDERSTANDING NATURAL LANGUAGE," COGNITIVE
NOPAD, "UNDERSTANDING NATURAL LANGUAGE," COGNITIVE
NOPAD, "UNDERSTANDING NATURAL LANGUAGE," COGNITIVE
NOPAD, "UNDERSTANDING NATURAL LANGUAGE," COGNITIVE
NOPAD, "UNDERSTANDING NATURAL LANGUAGE," COGNITIVE
NOPAD, "UNDERSTANDING NATURAL LANGUAGE," COGNITIVE
S, MATHEMATICAL STRUCTURE OF LANGUAGE, INTERSCIENCE
N OF INFORMATION PROCESSING CONGRESS, NORTH HOLLAND
N OF INFORMATION PROCESSING CONGRESS, NORTH HOLLAND
N OF INFORMATION PROCESSING CONGRESS, NORTH HOLLAND
N WONDERLAND AND THROUGH THE LOOKING GLASS, ALMOND
TH, THE ART OF COMPUTER PROGRAMMING, ADISON-WESLEY
A. KAPLAN, THE CONDUCT OF INQUIRY, CHANDLER
RATION PROCESSING, THE MIT PRESS, CAMBRIDGE. = M.R.
PROGRAM AND THEORY OF LANGUAGE," COMMUNICATION+
ELEMENTS DE SYNTAXE STRUCTURALE, PARIS, KLENCHSK+
R NATURAL LANGUAGE," STATISTICAL METHODS IN LINGUI+
TED AT SHARE XXVII, NEW YORK CITY, AUGUST 12, 197+
FFIC AND TECHNICAL LITERATURE, WILEY-INTERSCIENCE,+
ISH," PROC. OF FALL JOINT COMPUTER+ O.C. CLARKE AND
ISH," PROC. OF FALL JOINT COMPUTER+ O.C. CLARKE AND
DAY-AMERICAN ENGLISH, [BROWN UNIVERSITY, PROVIDENCE,
WE, "DIRECT+ACCESS MEMORY RETRIEVAL USING TRUNCATED
OOD FOR REPRESENTATION OF SYNTHETIC AND ANALYTICAL

- [illegible]

- COMMUNICATIONS OF THE ACM, 5(12), 103-144(1962). " G. SALTON, "MANIPULATION OF TREES IN INFORMATION RETRIEVAL," COM 0162
SCHWARZ, "A COMPUTATIONAL MODEL OF VERBAL UNDERSTANDING," PRO 0124
C OF THE FALL JOINT C. R. SIMMONS, J. BURGER AND R. SCHWARZ, "A COMPUTATIONAL MODEL OF VERBAL UNDERSTANDING," PRO 0014
STETH, "RELATIONAL LANGUAGES," REPORT TO NATIONAL SCIENCE FOUNDATION OFFICE OF SCIENCE INFORMATION SERVICE, O-0014
" REPORT TO NATIONAL SCIENCE FOUNDATION OFFICE OF SCIENCE INFORMATION SERVICE, O-S-U, COLUMBUS, OHIO, 1970. " 0014
ANALYSIS PROCEDURES IN THE SCIENCES OF MAN," SOCIAL SCIENCE INFORMATION CENTER, CISC-TR-71-13, THE OHIO STATE UNIV, 0155
ON STORAGE AND RETRIEVAL," COMPUTER AND INFORMATION SCIENCE RESEARCH CENTER, CISC-TR-71-13, THE OHIO STATE UNIV, 0160
" STORAGE AND RETRIEVAL," AMERICAN SOCIETY FOR INFORMATION SCIENCE 23(13), 195-219 (1972). " OMEY, "LINGUISTICS AND INF 0014
219. C.A. MONTGOMERY, "LINGUISTICS AND INFORMATION SCIENCES OF MAN," SOCIAL SCIENCE INFORMATION SERVICE, 8(1), 17-45(1961) 0155
97). " J. GADOM, "SEMANTIC ANALYSIS PROCEDURES IN THE SCIENTIFIC AND TECHNICAL LITERATURE, WILEY-INTERSCIENCE, NEM, 0137
MAIZELL, J.F. SMITH AND T.E. SINGER, ABSTRACTING SEEN FROM A CONSIDERATION OF MODALITY AND MOD, IN ENGLISH," 0054
HALLIDAY, "FUNCTIONAL DIVERSITY IN LANGUAGE AS SEMANTIC INFORMATION PROCESSING, THE MIT PRESS, CAMBRIDGE, M 0054
QUILLIAN, "SEMANTIC MEMORY," IN MI MINSKY (ED.), SEMANTIC THEORY," LANGUAGE, 39(2), 1963. " 0166
J.J. KATZ AND J.A. FODOR, "THE STRUCTURE OF A SEMIGROUPS, ACADEMIC PRESS, NEW YORK, 1968. " N M-A. ARSIB (0003
ED.), ALGEBRAIC THEORY OF MACHINES, LANGUAGES, AND SEMIGROUPS, ACADEMIC PRESS, NEW YORK, 1968. PP. 313-327. " 0036
(ED.), ALGEBRAIC THEORY OF MACHINES, LANGUAGES, AND SENTENCE DIAGRAMMING," 1961 INTERNATIONAL CONFERENCE ON MACH 0166
THE TRANSLATION OF LANGUAGES," W. PLATH, "AUTOMATIC RETRIEVAL USING TRUNCATED RECORD NAMES," SOFTWARE AGE SEPTEMBER, 28-33(1967). " L. LOWE, "DIRECT-ACCESS MEMORY RET 0066
RETRIEVAL USING TRUNCATED RECORD NAMES," SOFTWARE AGE SERVICE, O-S-U, COLUMBUS, OHIO, 1970. " REPORT TO NATION 0014
AL SCIENCE FOUNDATION OFFICE OF SCIENCE INFORMATION SERVICE, PHILADELPHIA, PENNSYLVANIA, 1970. PP. 71-81. " ALLOQU 0095
ON INFORMATION RETRIEVAL, MEDICAL DOCUMENTATION SHANK AND L. TESLER, "A CONCEPTUAL DEPENDENCY PARSER FOR NAT 0164
LINGUISTICS," STATISTICAL METHODS IN LINGUI, R.C. MOHOLA - STRUCTURE AND IMPLEMENTATION, PRESENTED AT 0037
LINGUISTICS, COLING, PREPRINT NO. 2, FOCK, STO ENGLISH WORDS," JOURNAL OF THE ACM, 5, KLEIN AND R. SIMMONS, J. BURGER AND R. SCHWARZ, "A COMPUTATIONAL MODEL OF 0014
VERBAL UNDERSTANDING," PROC OF THE FALL JOINT C. R. SIMMONS, J. BURGER AND R. SCHWARZ, "A COMPUTATIONAL MODEL OF 0014
VERBAL UNDERSTANDING," PROC OF THE FALL JOINT C. R. QUILLIAN, "THE TEACHABLE LANGUAGE COMPREHENDER," Y-INTERSCIENCE, " R.E. MAIZELL, J.F. SMITH AND T.E. SINGER, ABSTRACTING SCIENTIFIC AND TECHNICAL LITERATURE, WILE 0137
TORAGE AND RETRIEVAL," JOURNAL OF CHEMICAL DOCS, H. TESLER, "JOURNAL OF CHEMICAL DOCUMENTATION 10, SMITH AND T.E. SINGER, ABSTRACTING SCIENTIFIC AND TECHNICAL L 0137
RETRIEVAL, WILEY-INTERSCIENCE. " R.E. MAIZELL, J.F. SNOBOL4 - STRUCTURE AND IMPLEMENTATION, PRESENTED AT SHARE XX 0037
XVI, NEW YORK CITY, AUGUST 12, 1970. R.E. GRISWOLD, SOCIAL SCIENCE INFORMATION, 8(1), 17-45(1969). " GARDIN, "S 0155
EMANTIC ANALYSIS PROCEDURES IN THE SCIENCES OF MAN," SOCIETY FOR INFORMATION SCIENCE 23(13), 195-219 (1972). " OME 0014
BY "LINGUISTICS AND INFORMATION SCIENCES," AMERICAN SOCIETY OF AMERICA AT THE OHIO STATE UNIVERSITY, 1970. " 0137
PRESENTED AT THE SUMMER INSTITUTE OF THE LINGUISTIC SOFTWARE AGE, SEPTEMBER, 28-33(1967). " L. LOWE, "DIRECT-ACC 0066
ESS MEMORY RETRIEVAL USING TRUNCATED RECORD NAMES," SOME OBSERVATIONS," LANGUAGE, 40(4), 511-525(1964). " 0162
D.G. HAYES, "DEPENDENCY THEORY: A FORMALISM AND SOME OBSERVATIONS," LANGUAGE, 40(4), 511-525(1964). " 0161
D.G. HAYES, "DEPENDENCY THEORY: A FORMALISM AND SOME OBSERVATIONS," LANGUAGE, 40(4), 511-525(1964). " 0070
D.G. HAYES, "DEPENDENCY THEORY: A FORMALISM AND SPACE CO., PALO ALTO, CALIFORNIA, 1968 (AD 667 473). " RHAT 0070
INDEXING AND ABSTRACTING, LOCKHEED MISSILES AND SPACE CO., PALO ALTO, CALIFORNIA, 1968 (AD 667 473). " RHAT 0070
VE ABSTRACTING AND ABSTRACTING, LOCKHEED MISSILES AND SPACE CO., PALO ALTO, CALIFORNIA, 1968 (AD 667 473). " RHAT 0070
" L. OULAY, "AUTOMATIC DETERMINATION OF THE PARTS OF SPEECH IN ENGLISH," PROC. OF IEEE 15(7), 1029(1963). " AND J 0062
" L. OULAY, "AUTOMATIC DETERMINATION OF THE PARTS OF SPEECH IN ENGLISH," PROC. OF IEEE 15(7), 1029(1963). " AND J 0062
ATES OF ENGLISH TEXT," MASTERS THESIS FROM THE OHIO STATE UNIVERSITY, MARCH, 1973. " GENERATING STRUCTURAL SURROG 0160
TE OF THE LINGUISTIC SOCIETY OF AMERICA AT THE OHIO STATE UNIVERSITY, 1970. " PRESENTED AT THE SUMMER INSTITUTE 0137
M SCIENCE RESEARCH CENTER, CISC-TR-71-13, THE OHIO STATE UNIVERSITY, 1971. P.53. " SOME PROBLEMS FOR CASE GRAM 0137
MAR," WORKING PAPERS IN LINGUISTICS, NO.10, THE OHIO STATE UNIVERSITY, 1971. P.53. " SOME PROBLEMS FOR CASE GRAM 0137
N SCIENCE RESEARCH CENTER, CISC-TR-71-13, THE OHIO STATE UNIVERSITY, 1971. P.53. " SOME PROBLEMS FOR CASE GRAM 0137
M.W. RUTLEMAN, PRIVATE COMMUNICATION, THE OHIO STATE UNIVERSITY, 1973. " 0112
CONCEPTUAL DEPENDENCY PARSER FOR NATURAL LANGUAGE," STATISTICAL METHODS IN LINGUISTICS, 6, 31-51(1970). " A 0164
MUNO, W. STOLZ, P. TANNENBAUM AND P. CARSTENSEN, "A STOCHASTIC APPROACH TO THE GRAMMATICAL CODING OF ENGLISH," CO 0112
MUNO, W. STOLZ, P. TANNENBAUM AND P. CARSTENSEN, "A STOCHASTIC APPROACH TO THE GRAMMATICAL CODING OF ENGLISH," CO 0061
MUNO, W. STOLZ, P. TANNENBAUM AND P. CARSTENSEN, "A STOCHASTIC APPROACH TO THE GRAMMATICAL CODING OF ENGLISH," CO 0059
TATIONAL LINGUISTICS, COLING, PREPRINT NO. 2, FOCK, STOCHOLM, " AND TESLER, "CONCEPTUAL DEPENDENCY PARSER," COMPU 0137
H TO THE GRAMMATICAL CODING OF ENGLISH," COMMUN, W. STOLZ, P. TANNENBAUM AND P. CARSTENSEN, "A STOCHASTIC APPROACH 0061
H TO THE GRAMMATICAL CODING OF ENGLISH," COMMUN, W. STOLZ, P. TANNENBAUM AND P. CARSTENSEN, "A STOCHASTIC APPROACH 0061

- [illegible]

- [illegible]

- [illegible]

- JAN. 1972.** = PREPRINT FOR LANGUAGES AND LINGUISTICS
PRIES, THE STRUCTURE OF ENGLISH, MARCOURT, BRACE &
WORLD, INC., NEW YORK, 1952. = * GRAMMAR, LONDON, 1761, IN C.
0065
FRIES, THE STRUCTURE OF ENGLISH, MARCOURT, BRACE &
WORLD, INC., NEW YORK, 1952, P.70. =
C. 0070
PRIES, THE STRUCTURE OF ENGLISH, MARCOURT, BRACE &
WORLD, INC., NEW YORK, 1952, P.78. =
C. 0079
FRIES, THE STRUCTURE OF ENGLISH, MARCOURT, BRACE &
WORLD, INC., NEW YORK, 1952, P.74. =
C. 0084
XXVII, NEW YORK CITY, AUGUST 12, 1971. = * GRAISWALD, SHOBOLZ - STRUCTURE 0037
YORK CITY, AUGUST 12, 1971. = * GRAISWALD, SHOBOLZ - STRUCTURE 0037
YORK, 1930. =
L. BL 0003
YORK, 1930. P.5. =
L. BL 0049
YORK, 1932. = *H. GRAMMAR, LONDON, 1979 IN C. FRIES, THE STRUC 0036
YORK, 1932. = * GRAMMALS, LONDON, 1761, IN C. FRIES, THE STRUC 0009
YORK, 1932. P.70. =
C. FRIES, THE STRUC 0009
YORK, 1932. P.74. =
C. FRIES, THE STRUC 0066
YORK, 1932. P.75. =
C. FRIES, THE STRUC 0070
YORK, 1961. = H.A. GLEASON, AN INTRODUCTION TO DESER 0044
YORK, 1965. P.141. = *LICE'S ADVENTURES IN WONDERLAND AND THR 0069
YORK, 1965. = *C LINGUISTICS AND MACHINE TRANSLATION, IN AUTO 0065
YORK, 1968. = *E. IN UNIVERSALS IN LINGUISTIC THEORY, BACH, E 0134
YORK, 1968. = *IN UNIVERSALS IN LINGUISTIC THEORY, BACH, E 0134
YORK, 1968. = *H.A. ARABIS (ED.), ALGEBRAIC THEORY OF MACHI 0001
YORK, 1968. PP.313-327. = *B (ED.), ALGEBRAIC THEORY OF MACHI 0036
YORK, 1969. P.65. = W.A. COOK, S.J., INTRODUCTORY TO YAGHEM 0036
YORK, 1969. P.65. = W.A. COOK, S.J., INTRODUCTORY TO YAGHEM 0036
YORK, 1971. = *SMITH AND T.F. SIMMER, ABSTRACTING SCIENTIFI 0137
YOUNG, "GRAMMATICAL CLASS ASSIGNMENT BASED ON FUNCTION WORDS, 0095
ZEALAND, 1968." =
0004
- ZUCKY, J. FRIEDMAN, D.C. HALL AND D.E. WALKER, "THE MITRE SY 0053
ZUCKY, J. FRIEDMAN, D.C. HALL AND D.E. WALKER, "THE MITRE SY 0052
ZUCKY, J. FRIEDMAN, D.C. HALL AND D.E. WALKER, "THE MITRE SY 0113
0971). = AUTOMATIC INDEXING AND ABSTRACTING, LOCKHEAD MISSILES 0070
1(11), 176-200(1961). = INTERNATIONAL CONFERENCE ON MACHINE TRAN 0166
1-191(1972). = T. WINOGRAD, "NUMERS 0137
1-191(1972). = T. WINOGRAD, "NUMERS 0012
1-191(1972). = T. WINOGRAD, "NUMBER 0095
1-191(1972). = T. WINOGRAD, "NUMBERS 0034
10(2), 81-84(1970). = *LYTERM INDEX: A NEW CONCEPT IN INFORM 0137
10(2), 81-84(1970). = *A NEW CMCFPT IN INFORMATION STORAGE 0157
10(1), 342(1963). = *MONS, "A COMPUTATIONAL APPROACH TO GRAMA 0090
1029(1963). = *AND J.L. DOLEY, "AUTOMATIC DETERMINATION OF TM 0042
103-144(1962). = *AND J.L. DOLEY, "AUTOMATIC DETERMINATION OF T 0043
103-144(1962). = G. SALTON, "MANIPULATION OF TREES IN INFOR 0137
103-144(1962). = G. SALTON, "MANIPULATION OF TREES IN INFOR 0162
103-144(1962). =
0163
106-11(1962). = *PATENT INDEXING BY CONCEPT COORDINATION 0137
12(1), 459-476(1969). = *E COMPREHENSORS A SIMULATION PROGRAM 0137
13(10), 501-606(1970). = *TRANSITION NETWORK GRAMMARS FOR NAT 0137
13(10), 501-606(1970). = *TRANSITION NETWORK GRAMMARS FOR NAT 0112
13(10), 501-606(1970). = *TRANSITION NETWORK GRAMMARS FOR NAT 0112
13(10), 501-606(1970). = *TRANSITION NETWORK GRAMMARS FOR NAT 0044
13(10), 501-606(1970). = *J.M. DOLEY, "AUTOMATIC DETERMINATIO 0043
13(1) 1029(1963). = *AND J.L. DOLEY, "AUTOMATIC DETERMINATIO 0062
13(1), 1029(1963). = *AND J.L. DOLEY, "AUTOMATIC DETERMINATIO 0062
13(1), 1029(1963). = N. SAGER AND M. SALKOFF, "SYNTACTIC ANAL 0054
13(1)-108(1969). = N. SAGER AND M. SALKOFF, "SYNTACTIC ANAL 0112
13RD NATIONAL MEETING OF THE AMERICAN CHEMICAL SOCIETY, CHE 0197
1644, P.219. = J. MINNELL, DISCUSS COURSE CONCERNING THE PRECEDE 0194
17-45(1969). = *GARDIN, "SEMANTIC ANALYSIS PROCEDURES IN THE 0155
176-200(1961). = *INTERNATIONAL CONFERENCE ON MACHINE TRANSLATION 0164

- [illegible]

